SURVIVAL ANALYSIS

KK.
TKM. 03/04.
HUD
K

TESIS

KOMPARASI METODE POHON REGRESI DENGAN METODE REGRESI COX UNTUK ANALISIS DATA SURVIVAL

(ANALISIS FAKTOR YANG MEMPENGARUHI LAMA PERAWATAN PENYAKIT DIARE PADA BALITA DI RUMAH SAKIT)

MILIK
PERPUSTAKAAN
UNIVERSITAS AIRLANGGA
SURABAYA

NIZLEL HUDA

PROGRAM PASCASARJANA
UNIVERSITAS AIRLANGGA
SURABAYA
2003

KOMPARASI METODE POHON REGRESI DENGAN METODE REGRESI COX UNTUK ANALISIS DATA SURVIVAL

(ANALISIS FAKTOR YANG MEMPENGARUHI LAMA PERAWATAN PENYAKIT DIARE PADA BALITA DI RUMAH SAKIT)

TESIS

Untuk Memperoleh Gelar Magister

Dalam Program Studi Ilmu Kesehatan Masyarakat

pada Program Pascasarjana Universitas Airlangga

Oleh :

NIZLEL HUDA
NIM 090114305 M

PROGRAM PASCASARJANA
UNIVERSITAS AIRLANGGA
SURABAYA
2003

LEMBAR PENGESAHAN

TESIS INI TELAH DISETUJUI

TANGGAL 1 AGUSTUS 2003

Oleh

Pembimbing/Ketua

Prof. H. Kuntoro, dr., MPH, Dr.PH

NIP. 130 531 767

Pembimbing

Dr. Hari Basuki N., dr., M.Kes

NIP. 132 015 148

Mengetahui

Ketua Program Studi Ilmu Kesehatan Masyarakat

Program Pascasarjana Universitas Airlangga

Prof. Dr. H. R. Soedibjo HP. dr., DTM

NIP. 130 359 279

Telah diuji pada

Hari Jum'at tanggal 1 Agustus 2003

PANITIA PENGUJI TESIS

Ketua : Dr. Sunarjo, dr., MS., M.Sc

Anggota : 1. Prof. H. Kuntoro, dr., MPH, Dr.PH

2. Dr. Hari Basuki N, dr., M.Kes

3. Prof. Dr. Sarmanu, drh., MS

4. Dr. Arif Wibowo, dr., MS

5. Dr. Subijanto Marto Sudarmo, dr., SpA(K)

UCAPAN TERIMA KASIH

Puji Syukur penulis panjatkan ke hadirat Allah yang Maha Pengasih lagi Maha Penyayang atas segala Rahmat dan Karunia-Nya sehingga tesis ini dapat diselesaikan. Tesis ini disusun sebagai salah satu syarat akademik untuk menyelesaikan studi pada Program Studi Ilmu Kesehatan Masyarakat Program Pascasarjana Universitas Airlangga.

Penulis menyadari sepenuhnya bahwa penulisan tesis ini dapat diselesaikan berkat bantuan dari berbagai pihak, baik secara langsung maupun tidak langsung. Oleh karena itu penulis mengucapkan terima kasih dan penghargaan yang setinggi-tingginya kepada yang terhormat Bapak Prof. H. Kuntoro, dr., MPH., Dr.PH selaku pembimbing ketua dan bapak Dr. Hari Basuki N. dr., M.Kes sebagai pembimbing yang dengan penuh perhatian telah memberikan dorongan, bimbingan dan saran kepada penulis.

Pada kesempatan ini penulis juga menyampaikan ucapan terima kasih kepada yang terhormat :

1. Rektor Universitas Airlangga, Bapak Prof. Dr. Med Puruhito, dr., atas kesempatan dan fasilitas yang diberikan kepada penulis untuk mengikuti program Magister di Universitas Airlangga.
2. Direktur Program Pascasarjana Universitas Airlangga, Bapak Prof. Dr. H. Muhammad Amin, dr., SpP atas kesempatan yang telah diberikan kepada penulis untuk menjadi mahasiswa Program Magister pada Program Pascasarjana Universitas Airlangga.

3. Rektor Universitas Jambi, Bapak Prof. Dr. Ir. Ali Rahman M.A Rachman, M.A., yang telah memberi izin serta membebaskan penulis dari tugas rutin sebagai tenaga akademik selama mengikuti pendidikan Program Magister di Universitas Airlangga.
4. Ketua Program Studi Ilmu Kesehatan Masyarakat Program Pascasarjana Universitas Airlangga, Bapak Prof. Dr. H. R. Soedibjo HP. dr., DTM yang telah banyak memberikan motivasi dan wawasan ilmiah sehubungan dengan penyelesaian studi.
5. Ketua Peminatan Biostatistik Program Studi Ilmu Kesehatan Masyarakat, Bapak Prof. H. Kuntoro, dr., MPH., Dr. PH yang selalu memberikan arahan dan motivasi sejak penulis diterima sebagai mahasiswa Program Magister sampai pada tahap penyelesaian studi.
6. Bapak Dr. Subijanto Marto Sudarmo, dr., SpA(K) yang telah membimbing dan memberi bantuan serta arahan pengambilan data tentang faktor yang mempengaruhi lama perawatan penyakit diare pada balita di Rumah Sakit Dr. Soetomo Surabaya.
7. Bapak-bapak dan Ibu-ibu dosen pada Program Pascasarjana Universitas Airlangga khususnya Bapak-bapak dan Ibu-ibu dosen pembina mata kuliah Biostatistika yang telah banyak memberikan bekal ilmu dalam bidang Biostatistik.
8. Seluruh staf dan karyawan di lingkungan Pascasarjana Universitas Airlangga yang telah memberikan bantuan demi kelancaran administrasi selama penulis mengikuti pendidikan.

9. Seluruh rekan peserta Program Pascasarjana khususnya rekan-rekan Peminatan Biostatistik angkatan tahun 2001 yang telah memberikan bantuan dan sumbangan pikiran serta semangat untuk belajar sehingga penulis dapat menyelesaikan studi.
10. Ibu Hj. Rosdiana dan Papa H. Zainal Arifin, kakak Zaznelyetti dan Zazfineldi, adik-adik Zazwinaldi, Zazwelnefi, Zazriwaldi, Zaznawfeldi dan Zazmul Harmi, serta ponakan Mala Azitha dan Rahmi Azitha yang telah memberikan semangat dan motivasi kepada penulis dalam menyelesaikan studi pada Program Magister di Universitas Airlangga.

Semoga Allah SWT memberikan balasan yang setimpal kepada semua pihak yang telah memberikan bantuan kepada penulis.

Penulis menyadari bahwa dalam penulisan tesis ini banyak kekurangan yang diakibatkan oleh keterbatasan dalam diri penulis. Oleh karena itu dengan segala kerendahan hati penulis mengharapkan kritik dan saran yang sifatnya membangun demi kesempurnaan tesis ini. Mudah-mudahan tesis ini dapat bermanfaat bagi pembaca semua terutama untuk perkembangan ilmu pengetahuan di bidang Kesehatan Masyarakat, khususnya Biostatistik dan Ilmu-ilmu yang terkait dengannya.

Surabaya, Akhir Juli 2003

Penulis

RINGKASAN

Model regresi yang diperoleh dari metode kuadrat terkecil biasa (*Ordinary Least Square*/OLS) merupakan model regresi yang menghasilkan penaksir linier yang tidak bias yang terbaik (*Best Linear Unbiased Estimator*/BLUE) jika dipenuhi asumsi klasik. Asumsi klasik tersebut adalah nonmultikolinieritas, homoskedastisitas, nonotokorelasi, nilai rata-rata galat (*error*) populasi pada model stokastik sama dengan nol, variabel independen adalah nonstokastik (nilai konstan pada setiap kali percobaan yang dilakukan secara berulang) dan galat (*error*) berdistribusi normal.

Data dalam bidang kesehatan sering tidak memenuhi asumsi klasik. Masalah tersebut dapat diatasi dengan metode regresi yang tidak lagi terikat pada berbagai asumsi seperti metode pohon regresi dan metode regresi Cox.

Penelitian ini merupakan kajian statistik dalam membandingkan hasil analisis data survival dengan menggunakan metode regresi pohon dan metode regresi Cox. Penelitian ini diaplikasikan pada data lama perawatan penyakit diare pada balita di Rumah Sakit Dr. Soetomo Surabaya dengan variabel dependen adalah lama rawat dan variabel independen yaitu umur, jenis kelamin, dehidrasi, penyakit penyerta, lama diare, muntah dan status gizi.

Banyak data dalam penelitian ini adalah sebanyak 586 data. 57 data merupakan data yang tersensor dan 529 data merupakan data yang tidak tersensor.

Berdasarkan hasil analisis metode pohon regresi, variabel independen yang mempengaruhi lama perawatan penyakit diare pada balita di rumah sakit adalah lama diare, penyakit penyerta, status gizi dan dehidrasi. Karakteristik lama perawatan penyakit diare pada balita di rumah sakit dapat dikelompokkan ke dalam 7 kelompok yaitu :

1. Balita dengan lama diare sebelum dibawa ke rumah sakit lebih dari 13,5 hari maka rata-rata lama perawatannya di rumah sakit selama 10,556 hari.
2. Balita dengan lama diare sebelum dibawa ke rumah sakit kurang atau sama dengan 1,5 hari dan menderita penyakit Bronkopnemoni, Ensefalitis, Ensefalopati, Anemia, Marasmus dan penyakit lain-lain (bedah) maka rata-rata lama perawatannya 5,5 hari.
3. Balita yang menderita penyakit Bronkopnemoni, Ensefalitis, Ensefalopati, Anemia, Marasmus dan penyakit lain-lain (bedah) dengan $1,5 <$ lama diare $\leq 13,5$ hari maka lama perawatan balita di rumah sakit adalah 9,2 hari.
4. Balita dengan lama diare balita sebelum dibawa ke rumah sakit kurang atau sama dengan 13,5 hari dan menderita penyakit Meningitis, Malnutrisi, Kwarshiorkor, Marasmikwarshiorkor, Kejang Demam, Rhinopharyngitis dan tidak mempunyai penyakit penyerta, status gizi lebih dari 58,45 dengan dehidrasi sedang, berat dan tidak ada dehidrasi maka rata-rata lama perwatan balita di rumah sakit selama 5,857 hari
5. Lama diare balita sebelum dibawa ke rumah sakit kurang atau sama dengan 13,5 hari dan menderita penyakit Meningitis, Malnutrisi, Kwarshiorkor, Marasmikwarshiorkor, Kejang Demam, Rhinopharyngitis dan balita yang tidak mempunyai penyakit penyerta, gizi lebih dari 58,45 dan mengalami dehidrasi ringan maka rata-rata lama perawatan balita di rumah sakit selama 11 hari.

6. Balita yang menderita penyakit Meningitis, Malnutrisi, Kwarshiorkor, Marasmikwarshiorkor, Kejang Demam, Rhinopharyngitis dan balita yang tidak mempunyai penyakit penyerta dengan status gizi lebih dari 58,45 dengan lama diare kurang atau sama dengan 6,5 hari maka rata-rata lama perawatan balita di rumah sakit adalah 3,444 hari.
7. Balita yang menderita penyakit Meningitis, Malnutrisi, Kwarshiorkor, Marasmikwarshiorkor, Kejang Demam, Rhinopharyngitis dan balita tersebut tidak mempunyai penyakit penyerta dengan status gizi lebih dari 58,45 dengan 6,5 < lama diare ≤ 13,5 hari maka rata-rata lama perawatan balita di rumah sakit adalah 4,625 hari.

Jika diperhatikan hasil metode pohon regresi pada kelompok 4 dan 5 , dimana untuk balita penderita diare dan mengalami dehidrasi ringan, lama rawatnya lebih lama dibandingkan dengan balita penderita diare yang mengalami dehidrasi sedang, berat atau tidak ada dehidrasi. Hal ini diduga akibat dari penyakit penyerta Marasmikwarshiorkor dan Rhinopharyngitis yang diderita oleh balita penderita diare.

Berdasarkan hasil metode regresi Cox, variabel independen yang mempengaruhi lama perawatan penyakit diare pada balita di rumah sakit adalah lama diare, penyakit penyerta dan status gizi. Untuk menduga lama perawatan diare balita ke-i di Rumah Sakit Dr. Soetomo digunakan model di bawah ini:

$$\widehat{h_i(t)} = \exp(-0{,}726\ \text{Bronkopnemoni}_{(i)} - 1{,}531\ \text{Ensefalitis}_{(i)} - 1{,}064\ \text{Ensefalopati}_{(i)} - 0{,}641\ \text{Penyakit Lain-lain (bedah)}_{(i)} - 0{,}086\ \text{Lama diare}_{(i)} + 0{,}008\ \text{Status gizi}_{(i)})\ \widehat{h_0(t)}$$

Berdasarkan model di atas, balita penderita diare yang menderita penyakit Bronkopnemoni, Ensefalitis, Ensefalopati atau penyakit lain-lain (bedah) mempunyai lama perawatan yang lebih lama di rumah sakit. Begitu juga balita penderita diare dimana balita tersebut mempunyai waktu yang lama sebelum dibawa ke rumah sakit maka lama perawatan balita tersebut lebih lama dibandingkan dengan balita penderita diare yang langsung dibawa ke rumah sakit. Semakin tinggi status gizi balita penderita diare semakin sedikit waktu yang diperlukan untuk perawatan balita tersebut di rumah sakit dan taksiran median waktu survival diperoleh pada balita penderita diare dengan lama perawatannya di rumah sakit selama 4 hari.

Kelebihan metode pohon regresi adalah metode ini dapat menduga langsung berapa lama perawatan penyakit diare pada balita di rumah sakit dan kelemahan dari metode pohon regresi adalah metode ini tidak dapat menentukan kategori mana dari data yang berskala nominal atau ordinal sangat berpengaruh di dalam model Kelebihan dari metode regresi Cox adalah dapat diketahui secara langsung tingkat signifikansi variabel yang berpengaruh di dalam model dan kelemahannya adalah metode ini tidak dapat secara langsung menentukan lama perawatan penyakit diare balita di rumah sakit.

Kesimpulan dari penelitian ini adalah variabel yang berpengaruh terhadap lama perawatan penyakit diare pada balita di Rumah Sakit Dr. Soetomo dengan menggunakan metode pohon regresi adalah variabel lama diare, penyakit penyerta, status gizi dan dehidrasi, sedangkan variabel yang berpengaruh terhadap lama

perawatan diare balita di rumah sakit dengan menggunakan metode regresi Cox adalah variabel penyakit penyerta, lama diare dan status gizi. Berdasarkan uraian di atas maka disarankan, jika ingin menduga secara langsung lama perawatan penyakit diare pada balita di rumah sakit dengan variabel berskala nominal atau ordinal berkategori sama dengan 2 (dua) maka gunakankanlah metode pohon regresi, tetapi jika ingin menduga lama perawatan penyakit diare pada balita di rumah sakit dengan menggunakan probabilitas dimana variabel berskala nominal atau ordinal berkategori lebih dari 2 (dua) maka gunakankanlah metode regresi Cox.

SUMMARY

Regression model obtained from the Ordinary Least Square (OLS) is a model that result in the Best Linear Unbiased Estimator (BLUE) if the classical assumptions are fulfilled. The classical assumptions are nonmulticollinearity, homoskedasticity, nonautocorrelation, in stchastic model the mean of error in the population is zero, the independent variable is nonstochastic (constant value in each repeated trial) and the error has normal distribution.

Health data often do not fulfill the classical assumption. Such problem may be overcome by applying regression method that is not bound to any assumptions, such as regression trees and Cox regression methods.

This study was a statistical investigation comparing results of survival data analysis by means of regression trees method and Cox regression method. This study was applied to the data on the length of hospitalization in underfives with diarrhea in Dr Soetomo Hospital, Surabaya. The dependent variable was the length of hospitalization and the independent variables were age, sex, dehydration, accompanying diseases, length of diarrhea, vomiting and nutritional status.

The number of data in this study was 586, 57 of which were censored data, and the remaining 529 were uncensored.

Based on analysis of regression trees method, the independent variables affecting length of hospitalization for underfives with diarrhea were length of diarrhea, accompanying diseases, nutritional status, and dehydration. Characteristics of the length of hospitalization for those underfives might be clssified into 7 groups as follows:

1. Underfives with pre-hospitalized length of diarrhea of more than 13.5 days had mean of hospitalization of 10.556 days.
2. Underfives with pre-hospitalized length of diarrhea of less than or equal to 1.5 days and suffered from bronchopneumonia, encephalitis, encephalopati, anemia, marasmus, and other (surgical) diseases had mean of hospitalization of 5.5 days.
3. Underfives with pre-hospitalized length of diarrhea of more than 1.5 days and less than and equal to 13.5 days and suffered from bronchopneumonia, encephalitis, encephalopati, anemia, marasmus, and other (surgical) diseases had mean of hospitalization of 9.2 days
4. Underfives with pre-hospitalized length of diarrhea of less than or equal to 13.5 days and suffered from meningitis, manutrition, kwarshiorkor, marasmikwashiorkor, febrile convulsion, rhinopharyngitis, no accompying diseases, nutritional status of more than 58.54, and had severe, moderate and no dehydration, had mean of hospitalization of 5.857 days.
5. Underfives with pre-hospitalized length of diarrhea of less than or equal to 13.5 days and suffered from meningitis, manutrition, kwarshiorkor, marasmicwashiorkor. febrile convulsion, rhinopharyngitis, no accompying diseases, nutritional status of more than 58.54, and had mild dehydration, had mean of hospitalization of 11 days.

6. Underfives with pre-hospitalized length of diarrhea of less than or equal to 6.5 days and suffered from meningitis, manutrition, kwarshiorkor, marasmikwashiorkor, febrile convulsion, rhinopharyngitis, no accompying diseases, and nutritional status of more than 58.54, had mean of hospitalization of 3.444 days.
7. Underfives with pre-hospitalized length of diarrhea of more than or 6.5 and less than or equal to 13.5 days and suffered from meningitis, manutrition, kwarshiorkor, marasmicwashiorkor, febrile convulsion, rhinopharyngitis, no accompying diseases, and nutritional status of more than 58.54, and had severe, moderate and no dehydration, had mean of hospitalization of 4.625 days.

Result of regression trees method in groups 4 and 5 can be seen, in those groups underfives with diarrhea and mild dehydration had a longer of hospitalzation than those with moderate, severe, or no dehydration. This might be due to the accompanying diseases, i.e., marasmicwarshiorkor and rhinopharyngitis, suffered by those infants.

Based on the result of Cox regression method, the independent variables affecting the length of hospitalization of underfives with diarrhea were length of diarrhea, accompanying diseases, and nutritional status. To estimate length of hospitalzation of ith underfives with diarrhea at Dr Soetomo Hospital, the following model was used :

$$\mathbf{h_i(t) = \exp(-0.726\ bronchopnemonia_{(i)} - 1.531\ enchephalitis_{(i)} - 1.064\ encephalopathy_{(i)} - 0{,}641\ other\ (surgical)\ diseases_{(i)} - 0.086\ length\ for\ diarrhea_{(i)} + 0.008\ nutritional\ status_{(i)})\ h_0(t).}$$

Using above model, it was found that underfives with diarrhea and bronchopnemonia, encephalitis, encephalopathy, or other (surgical) diseases had longer length of hospitalization. Underfives with diarrhea with longer pre hospitalization also had longer length of hospitalization compared to those directly taken to the hospital. The higher the nutritional status, the shorter the length of their hospitalization. Estimated median of survival was obtained from underfives with length of hospitalization of 4 days.

The advantage of regression trees method is that it can directly estimate the length of hospitalization of those underfives, but it cannot determine category in data with nominal or ordinal scale that had hight influence in the model. In contrast, Cox regression model can determine directly the significance of the affecting variables in the model, while it cannot directly determine the length of hospitalization for underfives with diarrhea.

In conclusion, variables affecting length of hospitalization for underfives with diarrhea in Dr Soetomo Hospital using regression trees method are length of diarrhea, accompanying diseases, nutritional status, and dehidration; while using Cox regression method, those variables were accompanying diseases, length of diarrhea, and nutritional status. It can be recommended that regression trees method is appropriate to make direct estimation on the length of hospitalization for underfives with diarrhea using variables with nominal or ordinal scale and with the category of 2

(two). On the other hand, Cox regression method is appropriate to make estimation on the length of hospitalization for underfives with diarrhea using probability, in which the variables have nominal or ordinal scale with the category of more than 2 (two).

ABSTRACT

Regression analysis is one statistical tests commonly used for multifactorial data analysis. Regression analysis is interesting to use theoretically as it is based on mathematics. In addition, this analysis can be employed to investigate and model relations between variables. Regression analysis can be applied in variousdisciplines, such as engineering, physics, economics, management, biological and social sciences.

Problem that most commonly found in health data is that they do not fulfill the classical assumptions. Such problem may be overcome by means of regression analysis that is not bound to any assumptions, such as regression trees method and Cox regression analysis.

This study was aimed to survival data analysis (analyze factors affecting the length of hospitalization of underwifes with diarrhea) using regression trees method and Cox regression method. Data used in this study were those on the length of hospitalization of underfives with diarrhea in Dr Soetomo Hospital. The number of data in this study was 586, 57 of which were censored data, and the remaining 529 uncensored.

The cross validated relative error was 0.922 $\pm$ 0.041, the resubstitution relative error was 0.696 and the complexity was 45.497, the independent variables affecting the length of hospitalization in underfives with diarrhea, as revealed by means of regression trees method, were length of diarrhea, accompanying disseases, nutritional status, and dehydration, while variables affecting the length of hospitalization in underfives with diarrhea, as indicated by means of Cox regression method of α = 0.05, were length of diarrhea, accompanying diseases, and nutritional status.

Keywords : *survival data analysis, regression trees method, Cox regression method.*

DAFTAR ISI

Halaman

SAMPUL DEPAN i
SAMPUL DALAM ii
PRASYARAT GELAR iii
PERSETUJUAN iv
PANITIA PENGUJI TESIS v
UCAPAN TERIMA KASIH vi
RINGKASAN ix
SUMMARY xii
ABSTRACT xv
DAFTAR ISI xvi
DAFTAR TABEL xviii
DAFTAR GAMBAR xix

BAB 1 PENDAHULUAN 1
1.1 Latar Belakang Masalah 1
1.2 Rumusan Masalah 4
1.3 Tujuan Penelitian 4
1.3.1 Tujuan Umum 4
1.3.2 Tujuan Khusus 4
1.4 Manfaat Penelitian 5

BAB 2 TINJAUAN PUSTAKA 6
2.1 Metode Regresi 6
2.2 Metode Pohon Regresi 6
2.2.1 Definisi 6
2.2.2 Pembentukan Pohon Regresi 7
2.2.3 Aturan Pemilahan 8
2.2.4 Aturan *Growing* dan Kriteria *Goodness-of-Split* $\phi(s,t)$ 10
2.2.5 Penentuan Ukuran Pohon yang Layak 12
2.2.6 Menentukan Pohon Terbaik 16
2.3 Metode Regresi Cox 18
2.3.1 Model Regresi 18
2.3.2 Pengujian Parameter 21
2.3.3 Taksiran Fungsi Survival dan Fungsi *Hazard* 22
2.4 Data Survival 24
2.4.1 Fungsi Survival 28
2.4.2 Fungsi *Hazard* 28
2.5 Diare 29

BAB 3 METODE PENELITIAN 31
3.1 Jenis Penelitian 31
3.2 Populasi dan Sampel 31
3.3 Variabel Penelitian dan Definisi Operasional 31
3.4 Prosedur Pengumpulan Data 33
3.5 Teknik Analisis Data 33
3.6 Tahapan Analisis Data 33

BAB 4 HASIL PENELITIAN 37
4.1 Data Penelitian 37
4.2 Analisis Data 42
4.2.1 Analisis Metode Pohon Regresi 42
4.2.2 Analisis Metode Regresi Cox 47

BAB 5 PEMBAHASAN 49
5.1 Pembahasan Metode Pohon Regresi 49
5.2 Pembahasan Metode Regresi Cox 52
5.3 Komparasi Metode Pohon Regresi dengan Metode Regresi Cox 55

BAB 6 KESIMPULAN DAN SARAN 57
6.1 Kesimpulan 57
6.2 Saran 57

DAFTAR PUSTAKA 59

LAMPIRAN 61

DAFTAR TABEL

Halaman

Tabel 3.1	Banyak Pemilahan yang Dapat Dilakukan	34
Tabel 4.1	Frekuensi Tingkat Dehidrasi Balita	37
Tabel 4.2	Frekuensi Penyakit Penyerta yang Diderita Balita	38
Tabel 4.3	Gambaran Umum Data Lama Rawat, Umur, Lama Diare dan Status Gizi Balita	38
Tabel 4.4	Tabulasi Silang Penyakit Penyerta, Dehidrasi dan Status Gizi	40
Tabel 4.5	Tabulasi Silang Penyakit Penyerta, Dehidrasi dan Lama Diare	40
Tabel 4.6	Persentase Penyakit Penyerta pada Balita Penderita Diare dengan Dehidrasi Ringan	41
Tabel 4.7	Taksiran Nilai Koefisien dan Standar Error dari variabel Independen	48
Tabel 5.1	Taksiran *Hazard Ratio* untuk satu Balita	53
Tabel 5.2	Taksiran Baseline Hazard Function, Survivor Function dan Cumulative Hazard Function Lama Perawatan Balita Penderita Diare di Rumah Sakit	54
Tabel 5.3	Perbedaan Hasil Analisis Metode Pohon Regresi dengan Metode Regresi Cox	56

DAFTAR GAMBAR

Halaman

Gambar 2.1	Contoh Pembentukan Pohon Regresi	9
Gambar 4.1	Diagram Garis Data Lama Perawatan Penyakit Diare pada Balita di Rumah sakit	39
Gambar 4.2	Pohon Regresi Awal	42
Gambar 4.3	Plot Kesalahan Relatif dan Jumlah Simpul Lama Perawatan Penyakit Diare pada Balita di Rumah Sakit	43
Gambar 4.4	Pohon Regresi Optimal Lama Perawatan Penyakit Diare pada Balita di Rumah Sakit Dr. Soetomo	45

BAB 1

PENDAHULUAN

BAB 1

PENDAHULUAN

1.1. Latar Belakang Masalah

Montgomery (1992) mengatakan bahwa analisis regresi merupakan salah satu uji statistik yang banyak digunakan untuk analisis data multifaktor. Analisis regresi menarik untuk digunakan secara teoritis karena didasari oleh matematika. Di samping itu analisis ini dapat digunakan untuk menyelidiki dan memodelkan hubungan antar variabel. Analisis regresi dapat diaplikasikan dalam berbagai bidang seperti bidang teknik, fisika, ekonomi, manajemen, ilmu biologi dan ilmu sosial.

Dalam suatu persamaan regresi terdapat dua macam variabel yaitu variabel dependen dan variabel independen. Variabel dependen adalah variabel yang nilainya bergantung dari nilai variabel independen dan nilai variabel dependen tersebut ditaksir berdasarkan pada nilai variabel independen tertentu. Persamaan regresi yang digunakan untuk membuat taksiran mengenai nilai variabel dependen disebut persamaan regresi estimasi. Persamaan regresi estimasi adalah suatu formula matematis yang menunjukkan hubungan keterkaitan antara satu atau beberapa variabel yang nilainya sudah diketahui dengan satu variabel yang nilainya belum diketahui (Algifari, 2000)

Model regresi yang diperoleh dari metode kuadrat terkecil biasa (*Ordinary Least Square*/OLS) merupakan model regresi yang menghasilkan penaksir linier yang tidak bias yang terbaik (*Best Linear Unbiased Estimator*/BLUE). Keadaan

ini akan terjadi jika dipenuhi beberapa asumsi yang disebut dengan asumsi klasik. Menurut Montgomery (1992) asumsi klasik dalam analisis regresi tersebut adalah nonmultikolinieritas, homoskedastisitas, nonotokorelasi, nilai rata-rata galat (*error*) populasi pada model stokastik sama dengan nol, variabel independen adalah nonstokastik (nilai konstan pada setiap kali percobaan yang dilakukan secara berulang) dan galat (*error*) berdistribusi normal.

Masalah yang sering ditemui pada data dalam bidang kesehatan terutama pada data survival adalah asumsi klasik tidak dipenuhi. Masalah tersebut dapat diatasi dengan metode regresi yang tidak lagi terikat pada berbagai asumsi. Pendekatan yang pernah dilakukan adalah regresi dengan metode pohon biner pada penyekatan ruang variabel penjelas untuk melihat adanya perbedaan dugaan respon (Breiman, 1993).

Pohon regresi (*regression tree*) adalah salah satu metode yang menggunakan pohon keputusan (*decision tree*). Pohon keputusan ini dibentuk dengan menggunakan algoritma penyekat rekursif (*recursive partitioning*). Penggunaan algoritma ini diawali dengan munculnya program *Automatic Interaction Detection* (AID) dari Morgan dan Sonquist pada tahun 1963 (David and Anderson, 1989). Metode pohon regresi ini dapat digunakan dalam kondisi data mempunyai variabel independen dari semua skala data yaitu skala nominal, ordinal, interval maupun skala rasio.

David dan Anderson (1989) mengemukakan bahwa kelebihan metode pohon regresi ini adalah dapat mengeksplorasi struktur data yang kompleks, mengidentifikasi variabel independen yang mempunyai hubungan struktural

ADLN - Perpustakaan Universitas Airlangga

dengan variabel dependen, memprediksi variabel dependen dari satu atau beberapa amatan lain, mendeteksi interaksi antar variabel secara lokal dan dapat menemukan subgrup data yang bermakna, interpretasi hasilnya lebih mudah dari pada persamaan regresi linear dan resisten terhadap pengaruh pencilan.

Metode regresi Cox merupakan salah satu metode regresi yang sering digunakan dalam bidang kesehatan untuk menganalisis data survival. Pada umumnya data survival ini merupakan data tersensor karena data tentang lama waktu sampai terjadi kesembuhan seseorang tidak teramati secara lengkap tetapi tersensor oleh waktu yang ditentukan sebelum kejadian tersebut timbul, salah satu contohnya adalah kasus perawatan penyakit diare pada balita.

Sudiyanto (dalam Purwito, 1990) menyatakan bahwa berdasarkan laporan statistik rumah sakit mengenai angka kesakitan dan kematian, anak-anak (balita) yang dirawat di bangsal anak lebih banyak menderita penyakit diare dari pada penyakit lainnya. Lamanya perawatan diare pada balita di rumah sakit dipengaruhi oleh beberapa faktor seperti umur, jenis kelamin, tingkat dehidrasi, penyakit penyerta yang diderita balita, lama diare sebelum balita dibawa ke rumah sakit, muntah dan status gizi balita.

Berdasarkan uraian di atas ingin dilakukan penelitian tentang komparasi metode pohon regresi dengan metode regresi Cox untuk menganalisis faktor yang mempengaruhi lama perawatan penyakit diare pada balita di rumah sakit.

1.2. Rumusan Masalah

Berdasarkan latar belakang masalah di atas, maka rumusan masalah dalam penelitian ini adalah :

1. Bagaimana hasil analisis faktor yang mempengaruhi lama perawatan penyakit diare pada balita di rumah sakit dengan menggunakan metode pohon regresi.
2. Bagaimana hasil analisis faktor yang mempengaruhi lama perawatan penyakit diare pada balita di rumah sakit dengan menggunakan metode regresi Cox.
3. Bagaimana perbandingan hasil analisis faktor yang mempengaruhi lama perawatan penyakit diare pada balita di rumah sakit dengan menggunakan metode pohon regresi dan metode regresi Cox.

1.3. Tujuan Penelitian

1.3.1 Tujuan Umum

Tujuan umum dari penelitian ini adalah untuk membandingkan hasil analisis data survival dengan menggunakan metode pohon regresi dan metode regresi Cox.

1.3.2 Tujuan Khusus

Tujuan khusus dari penelitian ini adalah :

1. Mengetahui hasil analisis faktor yang mempengaruhi lama perawatan penyakit diare pada balita di rumah sakit dengan menggunakan metode pohon regresi.

2. Mengetahui hasil analisis faktor yang mempengaruhi lama perawatan penyakit diare pada balita di rumah sakit dengan menggunakan metode regresi Cox.
3. Membandingkan hasil analisis faktor yang mempengaruhi lama perawatan penyakit diare pada balita di rumah sakit dengan menggunakan metode pohon regresi dan metode regresi Cox.

1.4. Manfaat Penelitian

Manfaat praktis dari penelitian ini adalah metode pohon regresi dapat dijadikan sebagai suatu metode alternatif untuk menganalisis data survival.

BAB 2

TINJAUAN PUSTAKA

BAB 2

TINJAUAN PUSTAKA

2.1 Metode Regresi

Sering terdapat dua atau lebih elemen pembentuk sistem yang saling berhubungan, hal ini terjadi karena pada dasarnya suatu sistem terdiri dari elemen yang saling tergantung satu sama lain dan bekerja untuk mencapai tujuan. Dalam masalah nyata yang dihadapi terdapat dua atau lebih variabel yang saling berhubungan satu sama lain dan hubungan tersebut dapat diselidiki lebih lanjut.

Secara umum jika satu variabel dependen (variabel respon) tergantung pada satu atau beberapa variabel independen (variabel penjelas) maka hubungan di antara variabel ini dicirikan melalui model statistik yang disebut model regresi atau persamaan regresi. Model regresi ini digunakan untuk meramalkan (memprediksi) sejauh mana perlakuan yang dicobakan berpengaruh terhadap variabel dependen yang diamati (Gaspersz, 1995).

2.2 Metode Pohon Regresi (*Regression Trees*)

2.2.1. Definisi

Pada persamaan regresi linier, persamaan regresi menunjukkan bagaimana hubungan antara variabel independen dengan variabel dependen. Hal ini juga berlaku pada pohon regresi. Variabel independen yang berpengaruh dalam regresi linier juga akan merupakan variabel independen yang berpengaruh di dalam pohon regresi.

Variabel yang berpengaruh di dalam pohon regresi adalah variabel yang menentukan adanya pemilahan (*splitting*) ruang variabel independen (Segal, 1988). Adanya pemilahan tersebut mengakibatkan perbedaan dugaan terhadap variabel dependen. Penggunaan metode pohon regresi lebih ditujukan untuk masalah prediksi dan metode tersebut bersifat nonparametrik karena tidak terikat pada berbagai asumsi.

2.2.2. Pembentukan Pohon Regresi

Berikut ini akan akan dijelaskan mengenai proses pembentukan pohon regresi dan kriteria serta ukuran yang digunakan. Sama seperti metode regresi biasa, metode pohon regresi mempunyai satu variabel dependen yang kontinyu Y dan p variabel independen $X_1, X_2, \ldots, X_p$. Menurut Segal (1988) pembentukan pohon regresi memerlukan empat komponen yaitu :

1. Suatu himpunan pertanyaan dikotomus dengan bentuk "Apakah $x_i \in A$?" dengan x_i merupakan suatu amatan sampel dan $A \subset X$ (X adalah ruang variabel independen). Jawaban dari pertanyaan tersebut menentukan sekatan atau split bagi ruang variabel independen. Amatan dengan jawaban ya masuk ke anak ruang A sedangkan amatan dengan jawaban tidak masuk ke ruang komplemen A. Anak ruang sampel yang terbentuk disebut simpul (*node*).
2. Kriteria *goodness-of-split* $\phi(s,t)$ yang merupakan alat evaluasi bagi pemilahan yang dilakukan oleh pemilah (*split*) s pada simpul t.

3. Ukuran yang digunakan untuk menentukan ukuran pohon yang layak (*right sized tree*).
4. Statistik yang digunakan sebagai ringkasan pada setiap akhir simpul. Pada setiap akhir simpul dari pohon regresi digunakan statistik rata-rata variabel independen sebagai dugaan variabel dependen.

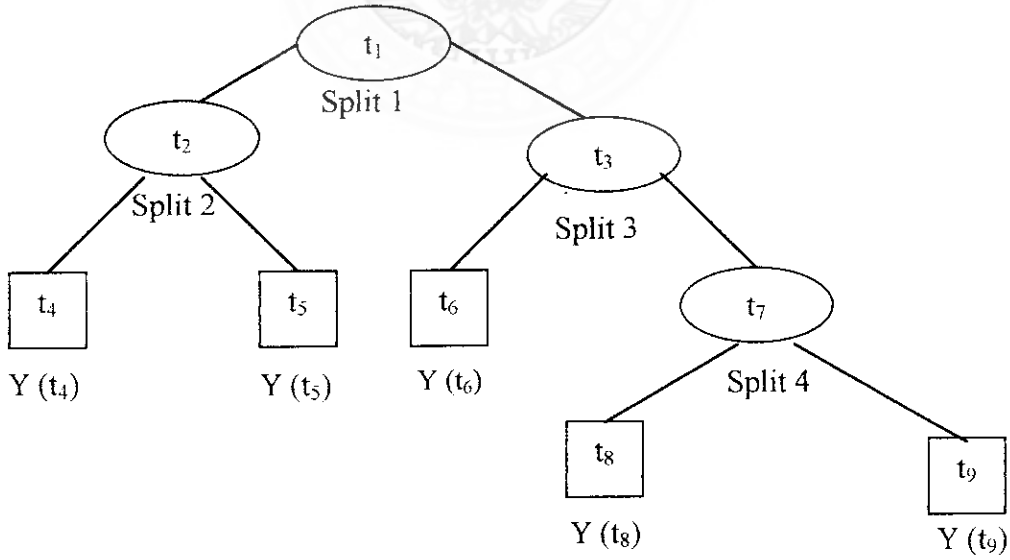
2.2.3. Aturan Pemilahan

Pohon regresi dibentuk melalui pemilahan data pada tiap simpul ke dalam dua simpul anak dengan aturan sebagai berikut :

1. Tiap pemilahan tergantung pada nilai yang hanya berasal dari satu variabel independen.
2. Untuk variabel kontinu X_j, pemilahan hanya berasal dari pertanyaan "Apakah nilai amatan variabel $X_j \leq c$?" untuk $c \in \Re^1$, c adalah bilangan terkecil dari kumpulan data amatan. Jika ruang sampelnya berukuran n dan terdapat sebanyak-banyaknya n nilai amatan berbeda pada variabel X_j maka akan terdapat sebanyak-banyaknya n–1 pemilah (*split*) berbeda yang dibentuk oleh gugus pertanyaan {"Apakah $X_j \leq c_i$?"} dengan i = 1, 2, ..., n-1 dan c_i adalah nilai tengah antara dua nilai amatan variabel X_j yang berurutan yang berbeda.
3. Untuk variabel independen kategorik, pemilahan yang terjadi berasal dari semua kemungkinan pemilahan berdasarkan terbentuknya dua anak gugus yang saling lepas (disjoint). Apabila variabel X_j merupakan variabel dengan

skala nominal yang kategori L maka akan ada 2^{L-1}-1 pemilahan, sedangkan jika variabel X_j tersebut berupa variabel dengan skala ordinal yang berkategori L maka ada L-1 pemilahan yang mungkin dilakukan.

Pada gambar 2.1 di bawah ini dapat dilihat sebuah contoh dari pembentukan pohon regresi. Penentuan simpul utama berdasarkan variabel terpenting. Pada simpul utama (*root node*) terdapat beberapa kelompok data di dalamnya. Selanjutnya simpul utama dipilah menjadi simpul anak kiri dan simpul anak kanan, demikian seterusnya sehingga diperoleh suatu simpul akhir (*terminal node*). Dalam gambar ini t_1 sebagai simpul utama (*root node*), t_2, t_3, t_7 sebagai simpul dalam (*internal nodes*), t_4, t_5, t_6, t_8, t_9 sebagai himpunan bagian terminal (*terminal set*), sedangkan y(t) merupakan hasil prediksi yang bernilai konstan.



Gambar 2.1 Contoh Pembentukan Pohon regresi

2.2.4. Aturan *Growing* dan Kriteria *Goodness-of-Split* $\phi(s,t)$

Pohon regresi dibentuk berdasarkan pemilahan rekursif berdasarkan kriteria tertentu. Proses pemilahan dilakukan pada tiap simpul dengan cara sebagai berikut :

1. Mencari semua kemungkinan pemilahan pada tiap variabel independen.
2. Memilih "pemilahan terbaik" masing-masing dari variabel independen dan memilih "pemilahan terbaik" dari "kumpulan pemilahan terbaik" tersebut. "Pemilahan terbaik" adalah pemilahan yang memaksimumkan ukuran kehomogenan di dalam masing-masing simpul anak relatif terhadap simpul induknya dan yang memaksimumkan ukuran pemisahan (*separation*) antara dua simpul anak kiri (t_L) dan simpul anak kanan (t_R).

Jumlah kuadrat sisaan digunakan sebagai kriteria kehomogenan pada tiap-tiap simpul. Misalnya, diketahui simpul t berisi sampel $\{(x_n,y_n)\}$, sedangkan n(t) adalah banyaknya amatan dalam simpul t dan rata-rata respon dalam simpul t adalah :

$$\bar{y}(t) = \frac{1}{n(t)} \sum_{x_n \in t} y_n$$

maka Jumlah Kuadrat Sisaan di dalam simpul t adalah :

$$JKS(t) = \sum_{x_n \in t} \left[y_n - \bar{y}(t) \right]^2$$

Misalkan ada pemilahan s yang menyekat t menjadi simpul anak kiri t_L dan simpul anak kanan t_R maka kriteria Jumlah Kuadrat Sisaan Terkecil adalah :

$$\phi(s,t) = JKS(t) - \{JKS(t_L) + JKS(t_R)\}$$

dan pemilah terbaik s^* adalah pemilah yang sedemikian sehingga

$$\phi(s^*,t) = \max_{s \in \Omega} \phi(s,t)$$

dengan Ω merupakan gugus yang berisi semua kemungkinan pemilahan.

Pohon regresi dibentuk melalui pemilahan simpul secara rekursif yaitu dengan memaksimumkan fungsi ϕ di atas. Pemilahan tersebut dapat dihentikan jika banyaknya amatan dalam simpul tersebut berjumlah "tertentu" atau pada saat ϕ lebih kecil dari suatu nilai ambang (*treshold*). Simpul yang terakhir dibentuk sebagai simpul akhir (*terminal node*) atau simpul daun (*leaf node*). Menurut Breiman (1993) banyaknya amatan pada simpul akhir kurang atau sama dengan 5. Dalam Program *Automatic Interaction Detection* (AID) dijelaskan bahwa suatu simpul akan dijadikan sebagai simpul akhir jika max $\phi(s,t) \leq 0{,}006\ JKS(t_1)$, t_1 adalah simpul utama (Breiman, 1993).

Penentuan nilai ambang bagi ϕ analog dengan penentuan nilai α bagi uji F pada penambahan variabel dalam regresi bertatar langkah maju (*forward stepwise regression*). Disini tidak ada lagi penambahan variabel bebas ke dalam model jika variabel tersebut memiliki taraf nyata uji F sebesar α (Therneau & Atkinson, 1997). Di samping itu Breiman (1993) memberikan sebuah kriteria alternatif yaitu Simpangan Mutlak Terkecil (*Least Absolute Deviation*) yang prosesnya sama dengan penggunaan Jumlah Kuadrat Terkecil.

ADLN - Perpustakaan Universitas Airlangga
MILIK
PERPUSTAKAAN
UNIVERSITAS AIRLANGGA
SURABAYA

2.2.5. Penentuan Ukuran Pohon yang Layak

Pohon yang dibentuk dengan aturan *splitting* dan aturan *growing* di atas berukuran sangat besar. Hal ini disebabkan karena aturan penghentian (*stopping rule*) yang digunakan hanya berdasarkan banyaknya amatan pada simpul akhir atau besarnya peningkatan tingkat kehomogenan.

Menurut Breiman (1993) lebih banyak split yang dilakukan mengakibatkan makin kecilnya tingkat kesalahan prediksi. Ini terjadi karena simpul akhir bisa hanya berisi satu amatan. Untuk menentukan ukuran pohon yang layak, masalah ini analog dengan regresi linear bertatar (*stepwise*) dimana R^2 terus meningkat bersamaan dengan banyak variabel independen yang masuk dalam model.

Pohon yang besar atau maksimal dapat menimbulkan dugaan overfitting dan sebaliknya kasus underfitting terjadi karena tidak adanya pemilahan lebih lanjut akibat adanya tetapan ambang $\phi(s^*,t)$, padahal pemilahan yang terjadi adalah layak. Masalah ini dapat diatasi dengan menentukan ukuran pohon yang layak. Untuk itu perlu dilakukan langkah-langkah sebagai berikut :

1. Penentuan pohon awal yang besar
2. Secara iteratif pohon tersebut dipangkas (*pruning*) menjadi deretan pohon yang makin kecil dan tersarang.
3. Pemilihan pohon terbaik dari deretan ini digunakan uji sampel (*test sample*) atau sampel validasi silang (*cross validation sample*)

Menurut Breiman (1993) pemangkasan dilakukan dengan menggunakan ukuran *cost complexity* minimum. Untuk sembarang pohon T yang merupakan subpohon dari pohon terbesar T_{max} diperoleh ukuran *complexity* $\left|\tilde{T}\right|$. Ukuran *complecixity* tersebut adalah banyaknya simpul akhir. Dalam regresi biasa, ukuran *complexity* tersebut sama dengan derjat bebas model. Selanjutnya Breiman (1993) menjelaskan, untuk suatu $\alpha \geq 0, \alpha \in \Re^1$ ukuran *cost complexity* adalah :

$$R_\alpha(t) = R(T) + \alpha\left|\tilde{T}\right|$$

di mana $R_\alpha(t)$ adalah *cost complexity*, $R(T)$ adalah *resubstitution estimate* yang digunakan yaitu Jumlah Kuadrat Sisaan atau Jumlah Simpangan Mutlak, α adalah parameter *complexity* mengenai cost bagi penambahan satu simpul akhir pada pohon T dan $\left|\tilde{T}\right|$ = ukuran *complexity*.

Lebih jauh Breiman (1993) menjelaskan bahwa langkah awal pemangkasan dilakukan terhadap T_1 yaitu suatu subpohon yang memenuhi kriteria $R(T_1)=R(T_{max})$. T_1 diperoleh dari T_{max} dengan cara mengambil t_L dan t_R yang merupakan simpul anak kiri dan simpul anak kanan dari T_{max} yang dihasilkan dari simpul induk t. Jika diperoleh $R(t) = R(t_l) + R(t_R)$ maka pangkaslah simpul anak t_L dan t_R tersebut. Ulangi proses ini sampai tidak ada lagi pemangkasan yang mungkin. Hasilnya adalah pohon T_1 yang memenuhi kriteria di atas.

Inti dari pemangkasan *cost complexity* minimum adalah pemotongan *weakest-link*. Untuk sembarang T_t yang merupakan cabang (*branch*) dari T_1, didefinisikan :

$$R(T_t) = \sum_{t \in \tilde{T}_t} R(t)$$

dengan $\tilde{T}_t$ adalah gugus simpul akhir dari T_t. Untuk sembarang simpul dalam (*internal node* atau *nonterminal node*) t dari pohon T_1 berlaku sifat $R(t)>R(T_t)$. {t} adalah subcabang dari T_t yang hanya terdiri dari satu simpul. Ukuran *cost complexity* dari subcabang{t} adalah :

$$R_\alpha(\{t\}) = R(t) + \alpha$$

dan ukuran *cost complexity* dari cabang T_t adalah :

$$R_\alpha(T_t) = R(T_t) + \alpha\left|\tilde{T}_t\right|$$

Oleh karena T_t merupakan cabang dan {t} merupakan subcabang yang terdiri dari satu simpul dalam maka :

$$R_\alpha(T_t) < R_\alpha(\{t\})$$

Ini berarti *cost complexity* cabang T_t lebih kecil daripada *cost complexity* subcabang {t}. Pada nilai kritis α, nilai *cost complexity* cabang T_t sama dengan nilai *cost complexity* subcabang {t}. Besar nilai kritis α tersebut adalah :

$$\alpha = \frac{R(t) - R(T_t)}{\left|\tilde{T}_t\right| - 1} \qquad (2.1)$$

untuk $t \in T_1$ definisi fungsi $g_1(t)$:

$$g_1(t) = \begin{cases} \dfrac{R(t) - R(T_t)}{\left|\tilde{T}_t\right| - 1}, t \notin \tilde{T}_1 \\ +\infty, t \in \tilde{T}_1 \end{cases}$$

Jika simpul tersebut memenuhi kriteria :

$$g_1\left(\bar{t}_1\right) = \min_{t \in T_1} g_1(t)$$

dan

$$\alpha_2 = g_1\left(\bar{t}_1\right)$$

maka $\bar{t}_1$ adalah *weakest-link* dalam T_1. Dengan demikian $\bar{t}_1$ adalah simpul dalam pertama yang membuat $R_\alpha(\{t\})=R_\alpha(T_t)$ dan mengakibatkan persamaan (2.1) terpenuhi, sedangkan α_2 adalah nilai dari parameter *complexity* dimana kesamaan tersebut terjadi. Selanjutnya ganti $\{\bar{t}_1\}$ tersebut dengan cabang $T_{\bar{t}_1}$, yaitu suatu cabang yang mempunyai simpul utama $\bar{t}_1$ dan pangkaslah cabang $T_{\bar{t}_1}$ tersebut dari pohon T_1. Hasil pemangkasannya adalah T_2 yaitu :

$$T_2 = T_1 - T_{\bar{t}_1}$$

Dengan demikian pohon T_2 adalah subpohon yang memenuhi kriteria *cost complexity* minimum dan nilai parameter *complexity* sebesar α_2.

Langkah selanjutnya adalah melakukan pemangkasan pada subpohon T_2 dengan proses sama seperti di atas. Jika terdapat *weakest-link* kembar pada

langkah ke-k, k=1, 2, ..., misalkan $g_k(\bar{t}_k) = g_k(\bar{t}'_k)$ maka pemangkasan yang dilakukan adalah :

$$T_{k-1} = T_k - T_{\bar{t}_k} - T_{\bar{t}'_k}$$

Hasil yang diperoleh dari pemangkasan di atas, berupa barisan subpohon yang tersarang dan makin lama makin kecil, yaitu $\{T_1, T_2, \ldots, \{t\}\}$ dengan $T_1 > T_2 > \ldots > \{t\}$. Ini dibaca : pohon T_1 adalah induk bagi pohon T_2, pohon T_2 adalah induk bagi pohon T_3, demikian seterusnya. Barisan α dalam urutan meningkat yaitu $\{\alpha_1, \alpha_2, \alpha_3 \ldots\}$ dengan $\alpha_1 = 0$, $\alpha_2 > \alpha_1$ demikian seterusnya.

2.2.6. Menentukan Pohon terbaik

Menentukan pohon terbaik dengan menggunakan *resubstitution estimate* R(T), cendrung dipilih pohon yang terbesar T_1 karena semakin besar pohon maka R(T) semakin kecil. Menurut Breiman (1993) ada dua estimate bagi R(t) yaitu :

1. *Test Sample Estimates* atau $R^{ts}(T)$
2. *Cross-validation Estimates* atau $R^{CV}(R)$

Untuk mendapatkan *test sample estimate*, data amatan dibagi menjadi dua secara acak menjadi *learning sample* L_1 dan *testing sample* L_2. L_1 digunakan untuk membentuk deretan pohon $\{T_k\}$ dengan cara pemangkasan dan L_2 digunakan untuk untuk membentuk $R^{ts}\{T_k\}$. Jika L_2 berukuran n_2 maka

$$R^{ts}(T_k) = \frac{1}{n_2} \sum_{(x_n, y_n) \in L_2} \left[y_n - \hat{y}_k(x_n) \right]^2$$

dengan $\hat{y}(x_n)$ adalah dugaan respon (variabel dependen) dari amatan ke-n pada pohon ke-k. Adapun pohon terbaik menurut Breiman (1993) adalah T_{k_0} yang memenuhi kriteria :

$$R^{ts}(T_{k_0}) = \min_k R^{ts}(T_k)$$

Selanjutnya Breiman (1993) menjelaskan bahwa untuk membentuk *cross-validation estimate* dengan V-fold amatan induk L yang dibagi secara acak menjadi V kelompok yaitu L_1, L_2, ..., L_v yang berukuran sama. *Learning sample* ke-v adalah $L^{-v} = L - L_v$, v = 1, 2, ..., V yang digunakan untuk membentuk deretan pohon $\{T_k\}$dan deretan parameter *complexity* $\{\alpha_k\}$. Jadi terdapat v deretan $\{T_k\}$ dan v deretan $\{\alpha_k\}$. Untuk membentuk deretan $\{T_k\}$ dan $\{\alpha_k\}$digunakan amatan induk L. Jika $\hat{y}_k^{-v}(x_n)$ adalah dugaan variabel dependen dari amatan ke-n pada pohon yang bersesuaian dengan α_k' yang dibentuk oleh *learning sample* ke-v dimana $\alpha_k' = \sqrt{\alpha_k \alpha_{k+1}}$ maka

$$R^{cv}(T_k) = \frac{1}{n} \sum_{v=1}^{v} \sum_{(x_n, y_n) \in L_v} \left[y_n - \hat{y}_k^{-v}(x_n) \right]^2$$

dan pohon yang terbaik adalah T_{k0} yaitu :

$$R^{cv}(T_{k0}) = \min_k R^{cv}(T_k)$$

Cross Validation estimate dengan 10-fold menghasilkan *resubstitution estimate* yang paling kecil (Breiman, 1993).

2.3. Metode Regresi Cox

2.3.1 Model Regresi

Menurut Collet (1994), model *hazard* proposional merupakan model dasar pada data survival. Model ini dikemukakan oleh Cox pada tahun 1972 dan kemudian dikenal dengan model regresi Cox (Model Cox). Model ini merupakan model regresi non parametrik. Di dalam model Cox fungsi *hazard* bagi amatan dengan kovariat $\underline{x} = (x_1, x_2, \ldots, x_p)$ dari variabel independen X_1, $X_2, \ldots, X_p$ adalah :

$$h_i(t) = \exp(\beta_1 x_{1i} + \beta_2 x_{2i} + \ldots + \beta_p x_{pi})\, h_0(t) \qquad (2.2)$$

atau

$$h(t,\underline{x}) = \exp(\underline{\beta}' \underline{x})\, h_0(t)$$

di mana $\underline{\beta}$ adalah vektor koefisien berdimensi-p dan $h_0(t)$ adalah *baseline hazard function*. Model di atas dapat ditulis sebagai :

$$\log\left\{\frac{h_i(t)}{h_0(t)}\right\} = \beta x_{1i} + \beta_2 x_{2i} + \ldots + \beta_p x_{pi}$$

atau

$$\log\left\{\frac{h_i(t)}{h_0(t)}\right\} = \underline{\beta}' \underline{x}$$

Model regresi Cox dapat juga dianggap sebagai model linear untuk logaritma rasio *hazard*. Penyusunan model regresi Cox dapat dijelaskan sebagai berikut. Misalkan ada sejumlah n data dengan r waktu *survive* sebenarnya dan n-r adalah waktu *survive* tersensor. Waktu *survive* sebenarnya r diurutkan menurut besarnya, $t_{(1)} \leq t_{(2)} \leq \ldots \leq t_{(r)}$ sehingga $t_{(j)}$ adalah urutan waktu sembuh ke-j. Sekumpulan individu yang sembuh pada pada waktu ke $t_{(j)}$ dinotasikan dengan $R(t_{(j)})$ dan jumlah $R(t_{(j)})$ disebut *risk set*.

Langkah selanjutnya dalam penyusunan model regresi cox adalah menyusun fungsi *likelihood*. Misalkan probabilitas pengamatan ke-i sembuh pada waktu $t_{(j)}$ dengan $t_{(j)}$ adalah satu pengamatan dari sejumlah r waktu sembuh $t_{(1)}, t_{(2)}, \ldots, t_{(r)}$ dan vektor variabel independen $\mathbf{x}_{(j)}$ adalah :

P(pengamatan dengan variabel $\mathbf{x}_{(j)}$ sembuh pada $t_{(j)}$|satu pengamatan sembuh pada $t_{(j)}$)

Probabilitas di atas dapat dinyatakan dengan :

$$\frac{P(pengamatan\ dengan\ variabel\ x_{(j)}\ sembuh\ pada\ t_{(j)})}{P(satu\ pengamatan\ sembuh\ pada\ t_{(j)})} \qquad (2.3)$$

Pembilang persamaan (2.3) merupakan hazard individu dengan vektor independen $\mathbf{x}_{(j)}$ pada waktu $t_{(j)}$. Untuk pengamatan ke-i, hazard individu ini dapat dinyatakan dengan $h_i(t_{(j)})$ sedangkan penyebut persaman (2.3) adalah jumlah *hazard* seluruh pengamatan yang sembuh pada pada waktu $t_{(j)}$ yang merupakan pengamatan dengan indeks *l* dalam $R(t_{(j)})$. Dengan demikian persamaan (2.3) menjadi :

$$\frac{h_i(t_{(j)})}{\sum_{l \in R(t_{(j)})} h_i(t_{(j)})}$$

Dengan demikian diperoleh :

$$\frac{\exp(\beta' x_{(j)})}{\sum_{l \in R(t_{(j)})} \exp(\beta' x_l)} \tag{2.4}$$

Selanjutnya Collett (1994) menyatakan bahwa fungsi likelihood untuk regresi Cox adalah :

$$L(\beta) = \prod_{j=1}^{r} \frac{\exp(\beta' x_{(j)})}{\sum_{l \in R(t_{(j)})} \exp(\beta' x_l)} \tag{2.5}$$

di mana $x_{(j)}$ adalah vektor kovariat dari amatan dengan waktu survival t_j, $R(t_j)$ adalah gugus amatan yang beresiko mati pada saat t_j dan l adalah panjang selang waktu kematian.

Jika ada n buah observasi waktu survival yaitu $t_1, t_2, \ldots, t_n$ dan δ_i adalah indikator sensor yang bernilai 0 (nol) jika data tersensor dan dan bernilai 1 (satu) jika data lengkap maka persamaannya menjadi :

$$L(\beta) = \prod_{i=1}^{n} \left(\frac{\exp(\beta' x_i)}{\sum_{j \in R(t_i)} \exp(\beta' x_l)} \right)^{\delta_i} \tag{2.6}$$

dan fungsi log-likelihoodnya adalah :

$$\log L(\beta) = \sum_{i=1}^{n} \delta_i \left\{ \beta' x_i - \log \sum_{l \in R(t_i)} \exp(\beta' x_l) \right\} \tag{2.7}$$

Pemaksimuman fungsi persamaan (2.7) dapat dilakukan dengan prosedur Nepton-Raphson.

2.3.2 Pengujian Paremeter

Ada dua macam pengujian parameter yaitu :

1. Uji parameter secara serentak.

   Untuk meguji parameter secara serentak, hipotesisnya adalah :

   $H_0 : \beta_1 = \beta_2 = \ldots = \beta_p$

   H_1 : Minimal ada satu $\beta_j \neq 0$, j = 1,2, …, p

   Statistik uji yang digunakan adalah uji Wald dengan rumus :

$$X^2_{hitung} = \hat{\beta}\, I(\hat{\beta})\, \hat{\beta}$$

   dengan

   β adalah vektor dari β_j, j = 1, 2, …, p

   $\hat{\beta}$ adalah dugaan dari β

   $I(\hat{\beta})$ adalah matriks informasi yang diamati

   Elemen (j,k) dari $I(\hat{\beta})$ adalah :

$$I(\hat{\beta}) = -\left\{ \frac{\partial^2 \log L(\beta)}{\partial \beta_j \;\; \partial \beta_k} \right\}$$

   dengan j = 1, 2, …, p dan k = 1, 2, …, p

   Kriteria uji adalah H_0 ditolak jika $X^2_{hitung} > X^2_{(p,\alpha)}$.

2. Uji parameter secara individual.

   Untuk menguji parameter secara individual, hipotesisnya adalah :

$H_0 : \beta_j = 0, j = 1, 2, \ldots, p$

$H_1 : \beta_j \neq 0, j = 1,2, \ldots, p$

Statistik uji yang digunakan adalah uji Wald dengan rumus :

$$X^2_{hitung} = \hat{\beta}^2 \, i(\hat{\beta})$$

dengan

$\hat{\beta}$ adalah taksiran dari β

$i(\hat{\beta})$ adalah matriks yang diamati

Elemen (j,k) dari $I(\hat{\beta})$ adalah :

$$i(\hat{\beta}) = -\left\{\frac{\partial^2 \log L(\beta)}{\partial \beta^2}\right\}$$

Kriteria uji adalah H_0 ditolak jika $X^2_{hitung} > X^2_{(1,\alpha)}$.

2.3.3 Taksiran Fungsi Survival dan Fungsi *Hazard*

Setelah diperoleh model regresi cox maka diperoleh pula taksiran fungsi survival dan fungsi *hazard*. Menurut Collett (1994), jika komponen linear dari model regresi cox terdiri dari p variabel independen $X_1, X_2, \ldots, X_p$ dan taksiran parameter dari variabel terebut adalah $\hat{\beta}_1, \hat{\beta}_2, \ldots, \hat{\beta}_p$ maka taksiran fungsi *hazard* pengamatan ke-i adalah :

$$\hat{h}_i(t) = \exp(\hat{\beta}' x_i) h_0(t)$$

ADLN - Perpustakaan Universitas Airlangga

MILIK
PERPUSTAKAAN
UNIVERSITAS AIRLANGGA
SURABAYA

dengan

i = 1, 2, …, n

x_i adalah vektor dari nilai variabel independen untuk individu ke-i.

$\hat{h}_0(t)$ adalah penduga *baseline hazard function.*

Taksiran *baseline hazard function* yang diperoleh dengan menggunakan pendekatan maximum likelihood adalah :

$$\hat{h}_0(t_{(j)}) = 1 - \hat{\xi}_j$$

Jika diasumsikan tidak terdapat *tied survival times* maka diperoleh :

$$\hat{\xi}_j = \left(1 - \frac{\exp(\hat{\beta}' x_{(j)})}{\sum_{l \in R(t_{(j)})} \exp(\hat{\beta}' x_l)}\right)^{\exp(-\hat{\beta}' x_{(j)})}$$

dengan

$x_{(j)}$ adalah vektor dari variabel independen

j = 1, 2, …, r

Selanjutnya Collett (1994) mengatakan bahwa taksiran *baseline survivor function* untuk $t_{(k)} < t < t_{(k+1)}$, k = 1, 2, …, r-1 adalah :

$$\hat{S}_0(t) = \prod_{j=1}^{k} \hat{\xi}_j \tag{2.8}$$

sedangkan taksiran *cumulative baseline hazard function* adalah :

$$\hat{H}_0(t) = -\log \hat{S}_0(t) = -\sum_{j=1}^{k} \log \hat{\xi}_j \tag{2.9}$$

Berdasarkan persamaan (2.8) dan (2.9) untuk $t_k \le t \le t_{(k+1)}$, k = 1, 2, ..., r-1 diperoleh taksiran *survivor function* dan pengamatan ke-i adalah :

$$\hat{H}_i(t) = \exp(\hat{\beta}' x_i)\hat{H}_0(t)$$

dan taksiran *cumulative hazard function* pengamatan ke-i adalah :

$$\hat{S}_i(t) = \left[\hat{S}_0(t)\right]^{\exp(\hat{\beta}' x_i)}$$

Jika ada *tied survival time* maka $\hat{\xi}_j^{\exp(\hat{\beta}' x_i)}$ dapat dinyatakan dengan $\exp\left(e^{\hat{\beta}' x_i} \log \hat{\xi}_j\right)$ dan $\exp\left(e^{\hat{\beta}' x_i} \log \hat{\xi}_j\right)$ dapat dinyatakan dengan $\exp\left(e^{\hat{\beta}' x_i} \log \hat{\xi}_j\right) \approx 1 + e^{\hat{\beta}' x_i} \log \hat{\xi}_j$, sehingga untuk $t_k \le t \le t_{(k+1)}$, k = 1, 2, ..., r-1 taksiran fungsi survival menjadi :

$$S_0^*(t) = \prod_{j=1}^{k} \exp\left(\frac{-d_j}{\sum_{l \in R(t_{(j)})} \exp(\hat{\beta}' x_l)}\right)$$

dan fungsi *hazard* menjadi :

$$H_0^*(t) = -\log S_0^*(t) = \sum_{j=1}^{k} \frac{d_j}{\sum_{l \in R(t_{(j)})} \exp(\hat{\beta}' x_l)}$$

2.4 Data Survival

Data survival adalah istilah yang digunakan bagi data tentang lamanya waktu sampai terjadinya suatu kejadian. Kejadian yang paling sederhana antara lain

adalah terjadinya suatu penyakit atau kambuhnya suatu penyakit. Analisis survival ini berhubungan dengan metode statistika.

Data survival biasanya tidak simetris yang mengakibatkan histogram data survival ini cenderung menceng ke kanan atau menceng ke kiri. Peneliti tidak dapat mengamati objek penelitian sampai timbulnya kejadian sehingga objek penelitian tersebut tersensor.

Menurut Miller dkk (1981) ada 3 (tiga) jenis penyensoran pada data survival yaitu :

1. Penyensoran jenis I.

   Jika t_c adalah waktu yang menyatakan batas akhir dari suatu penelitian yang telah ditetapkan sebelumnya dan $T_1, T_2, \ldots, T_n$ adalah suatu variabel acak maka terdapat $Y_1, Y_2, \ldots, Y_p$ menyatakan waktu survival yang diperoleh dari :

$$Y_i = \begin{cases} T_i, jika \quad T_i \le t_c \\ t_c, jika \quad t_c < T_i \end{cases}$$

   dengan fungsi distribusi Y adalah positif karena $P\{T > t_c\} > 0$ pada $y = t_c$

2. Penyensoran jenis II.

   Jika n menyatakan banyak data responden, maka langkah awal dari penyensoran jenis II adalah menentukan suatu nilai r ($r < n$). Penelitian akan dihentikan jika telah ditemukan r responden dari n responden yang telah menunjukkan tanda-tanda yang diteliti.

3. Penyensoran secara acak.

   Jika $T_1, T_2, \ldots, T_p$ adalah suatu variabel acak dan $C_1, C_2, \ldots, C_n$ adalah suatu data yang menyatakan waktu penyensoran yang berhubungan dengan T_i maka akan diperoleh pasangan data $(Y_1, \delta_1), (Y_2, \delta_2), \ldots, (Y_n, \delta_n)$, di mana $Y_i = \min(T_i, C_i)$.

$$\delta_i = I(T_i \le C_i)\begin{cases} 1, & jika \;\; T_i \le C_i, \;\; T_i \;\; tidak \;\; disensor \\ 0, & jika \;\; t_c < T_i, \;\; T_i \;\; disensor \end{cases}$$

   Keterangan :

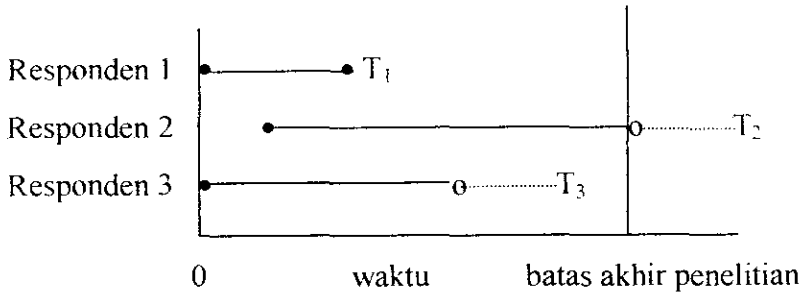
   $T_1, t_2, \ldots, t_p$ adalah waktu survival.

   $\delta_1, \delta_2, \ldots, \delta_p$ adalah informasi tentang penyensoran.

Penyensoran secara acak ini banyak digunakan pada penelitian klinis. Pada penelitian klinis sering terjadi responden mulai masuk dalam penelitian tidak dalam waktu yang bersamaan sehingga penyensoran secara acak ini sangat tepat untuk digunakan. Penyensoran juga terjadi karena :

1. Responden hilang atau mengundurkan diri dari penelitian.
2. Drop out, perlakuan yang diberikan kepada responden membahayakan, sehingga diputuskan untuk dihentikan.
3. Batas waktu penelitian telah berakhir.

Penyensoran secara acak tersebut dapat dilihat pada gambar 2.2 berikut ini.



Gambar 2.2 Penyensoran secara acak

Pada gambar di atas terlihat, responden 1 meninggal pada waktu T_1 sehingga merupakan pengamatan yang tidak sensor. Responden 2 masih hidup melewati batas akhir penelitian sehingga menghasilkan pengamatan sensor T_2^+. Responden 3 hilang dalam pengamatan yang juga merupakan pengamatan sensor T_3^+.

Menurut Collect (1994) ada dua fungsi yang menjadi pusat perhatian pada data survival yaitu fungsi survival (*function survivor*) dan fungsi *hazard* (*hazard function*). Waktu survival dari suatu individu (t) dapat dipandang sebagai nilai dari variable T. Perbedaan nilai T yang diambil mempunyai probabilitas distribusi dan T disebut variabel random dengan waktu survival. Jika variabel T mempunyai probabilitas distribusi yang didasari pada probabilitas fungsi density (*probability density function* atau f(t)), maka distribusi fungsi T adalah :

$$F(t) = P(T < t) = \int_0^t f(u)du$$

dengan probabilitas waktu survival kurang dari t.

2.4.1 Fungsi Survival

Menurut Lawless (1982), fungsi survival S(t) didefinisikan sebagai probabilitas bertahan hidup sampai waktu t ($t \geq 0$). S(t) adalah fungsi kontinyu monoton turun dengan S(0) = 1 dan $S(\infty) = \lim_{t \to \infty} it = 0$.

Hal di atas sesuai dengan pendapat Lee (1980) yang mengemukakan bahwa fungsi survival adalah peluang suatu individu dapat bertahan hidup lebih dari waktu t dan dinotasikan dengan S(t). Fungsi survival ini dapat diestimasi melalui proporsi individu yang hidup lebih dari t. Dengan demikian S(t) dapat ditulis sebagai :

$$S(t) = \frac{jumlah \quad individu \quad hidup > t}{jumlah \quad total \quad individu}$$

2.4.2 Fungsi Hazard

Fungsi *hazard* h(t) merupakan laju kegagalan atau kematian (kegagalan atau kematian sesaat) pada waktu t (Nelson, 1982), sedangkan h(t) Δt merupakan pendekatan probabilitas kegagalan atau kematian dalam interval [t, t+Δt] dengan syarat bertahan hidup sampai waktu t. Fungsi *hazard* h(t) mempunyai sifat-sifat :

$$h(t) \geq 0 \text{ untuk } \forall\ t \geq 0 \text{ dan } \int_0^\infty h(t)dt = \infty$$

Adapun hubungan antara fungsi *hazard* h(t) dengan fungsi survival S(t) adalah sebagai berikut :

$$h(t) = \frac{d[\log S(t)]}{dt}$$

sehingga

$$\log S(x)|_0^t = -\int_0^t h(x)dx$$

karena S(0) = 1, maka diperoleh :

$$S(t) = \exp\left(-\int_0^t h(x)dx\right).$$

Pada tahun 1986, Hougard berhasil menyusun suatu model probabilistik waktu survival untuk populasi heterogen dimana tiap-tiap individu dalam satu populasi diasumsikan mempunyai tingkat resiko kematian yang berbeda. Model Hougard ini menggunakan tiga parameter α, δ dan θ yang belum diketahui nilainya secara pasti, dengan fungsi hazard populasinya adalah :

$$h(t) = \delta(\theta + t))^{\alpha-1}, \quad 0 \le \alpha \le 1, \delta > 0, \theta \ge 0$$

dan fungsi daya tahan hidup populasinya adalah :

$$S(t) = \begin{cases} \left(\dfrac{\theta}{\theta + t}\right)^{\delta}, (\alpha = 0, \delta > 0, \theta \ge 0) \\ \exp\left\{-\dfrac{\delta}{\alpha}\left[(\theta + t)^{\alpha} - \theta^{\alpha}\right]\right\}, (0 < \alpha \le 1, \delta > 0, \theta > 0) \\ \exp\left(-\dfrac{\delta}{\alpha} t^{\alpha}\right), (0 < \alpha \le 1, \delta > 0, \theta = 0) \end{cases}$$

2.5 Diare

Menurut Depkes RI (2000) diare adalah suatu penyakit yang ditandai dengan adanya buang air besar lembek atau cair bahkan dapat berupa air saja yang frekuensinya lebih sering dari biasa (3 kali atau lebih dalam sehari). Kalangan

ADLN - Perpustakaan Universitas Airlangga

medis memberikan istilah penyakit diare dengan gastroenteritis. Penyakit diare yang sering terjadi disebabkan oleh toksin yang dihasilkan oleh mikroorganisme, misalnya V.cholera dan E. Coli.

WHO (1999) mengklasifikasikan lama diare menjadi 2 (dua) tipe yaitu diare akut dan diare persiten. Diare akut terjadi kurang dari 2 minggu sedangkan diare persiten berlangsung selama 2 (dua) minggu atau lebih.

Lubis (1992) mengemukakan bahwa diare pada balita di bawah 2 (dua) tahun disebabkan karena ibu tidak membersihkan dot, makanan tidak dipanaskan dan tidak mencuci tangan sebelum menyuapi balita. Selanjutnya menurut Krisna dan Supardi (1996), sejumlah faktor risiko yang mempengaruhi kejadian diare berdarah pada usia balita di Kabupaten Klaten adalah kebiasaan tidak mencuci tangan dan kebiasaan tak menutup hidangan.

Berdasarkan data di ruang Gastroenterologi anak RSUD Dr. Soetomo tahun 1998 terdapat 726 kasus diare akut yang meliputi 537 kasus diare pada usia bayi (14 orang meninggal) dan 189 kasus diare terjadi pada balita yang berusia 1 - 5 tahun. Selanjutnya berdasarkan Survey Kesehatan Rumah Tangga (SKRT) tahun 1995, kematian balita karena diare tercatat sebesar 35% (Ranuh, 2001).

BAB 3

METODE PENELITIAN

BAB 3

METODE PENELITIAN

3.1 Jenis Penelitian

Penelitian ini termasuk dalam penelitian terapan (applied research) dengan menerapkan metode pohon regresi dan metode regresi Cox pada data sekunder.

3.2 Populasi dan sampel

Populasi penelitian ini adalah semua balita yang menderita diare yang dirawat di bangsal anak RSUD Dr. Soetomo. Adapun sampel dalam penelitian ini adalah semua balita yang menderita diare yang dirawat di RSUD Dr. Soetomo dari bulan Juni 2002 sampai dengan bulan Mei 2003 yang berjumlah sebanyak 586 balita.

3.3 Variabel Penelitian dan Definisi Operasional

Sebagai variabel dependen dalam penelitian ini adalah lama perawatan balita yang menderita diare di rumah sakit (Y), sedangkan variabel independen adalah :

1. Umur (X_1).
2. Jenis kelamin (X_2).
3. Dehidrasi (X_3).
4. Penyakit penyerta (X_4).
5. Lama diare (X_5).
6. Muntah (X_6)

7. Status gizi (X_7)

Adapun definisi operasional dari variabel di atas adalah sebagai berikut :

1. Lama perawatan balita yang menderita diare adalah waktu yang diperlukan untuk merawat balita sejak masuk rumah sakit sampai balita tersebut pulang atau pindah kamar atau pulang paksa atau meninggal. Bagi balita yang pulang sesuai rekomendasi para medis dikatakan lama perawatannya lengkap, sedangkan bagi balita yang pulang paksa, pindah kamar dan meninggal maka lama perawatannya tersensor karena tidak teramati secara lengkap. Variabel ini diukur dalam hari dan data berskala rasio.
2. Umur adalah usia pada saat balita menderita diare. Variabel ini diukur dalam bulan dan data berskala rasio.
3. Jenis kelamin adalah jenis kelamin balita yang menderita diare. Variabel ini terdiri dari 2 (dua) kategori yaitu (1) laki-laki, (2) perempuan dan data berskala nominal.
4. Dehidrasi adalah tingkat dehidrasi yang dialami balita pada saat menderita diare. Variabel ini terdiri dari 4 kategori yaitu (1) dehidrasi ringan, (2) dehidrasi sedang, (3) dehidrasi berat, (4) tidak ada dehidrasi dan data berskala ordinal.
5. Penyakit penyerta adalah penyakit penyerta balita pada saat menderita diare. Variabel ini terdiri dari 14 kategori yaitu (1)Bronkopnemoni, (2)Meningitis, (3) Ensefalitis, (4) Meningoensefalitis, (5) Ensefalopati, (6) Malnutrisi, (7) Anemia, (8) Marasmus, (9) Kwarshiorkor, (10) Marasmikwarshiorkor, (11)

Kejang Demam, (12) Rhinopharyngitis, (13) Lain-lain (bedah), (14) Tidak ada. Data ini berskala nominal.

6. Lama diare adalah lama waktu balita mengalami diare sebelum dirawat di rumah sakit. Variabel ini diukur dalam jam dan data berskala rasio.
8. Muntah adalah adanya muntah yang meyertai diare pada balita. Variabel ini terdiri dari 2 kategori yaitu (1) ya, (2) tidak dan data berskala nominal.
9. Status gizi adalah keadaan gizi balita pada saat diare. Variabel ini diukur berdasarkan standar NCHS-WHO dan datanya berskala rasio.

3.4 Prosedur pengumpulan data

Data yang digunakan dalam penelitian ini adalah data sekunder yang diperoleh dari rekam medik penyakit diare di bagian anak Divisi Gastroenterologi RSUD Dr. Soetomo.

3.5 Teknik Analisis Data

Teknik analisis data yang digunakan adalah metode pohon regresi dan metode regresi Cox . Untuk metode pohon regresi digunakan program CART 4.0, sedangkan untuk metode regresi Cox digunakan program SPSS versi 10.

3.6 Tahapan Analisis Data

Tahap-tahap analisis data adalah sebagai berikut :

1. Eksplorasi data

2. Pembentukan model dengan pohon regresi melalui tahap-tahap sebagai berikut :

   a. Pembentukan pohon regresi dengan cara :

      i. Pemilihan pemilah dilakukan dengan cara rekursif atau dengan menggunakan kriteria tertentu. Pemilahan berdasarkan variabel umur, lama diare dan status gizi balita saat menderita diare di mana variabel-variabel tersebut merupakan variabel kontinyu dan banyak data untuk setiap variabel tersebut adalah 586 data, ada sebanyak 585 pemilahan yang mungkin dilakukan. Jika pemilahan berasal dari variabel jenis kelamin, tingkat dehidrasi, penyakit penyerta dan muntah maka banyak pemilahan yang dilakukan dapat dilihat pada tabel 3.1.

      Tabel 3.1 Banyak Pemilahan yang Dapat Dilakukan.

Variabel	**Skala Data**	**Banyak Kategori**	**Banyak Pemilahan**
Jenis kelamin	Nominal	2	1
Dehidrasi	Ordinal	4	3
Penyakit Penyerta	Nominal	14	16.383
Muntah	Nominal	2	1

      ii. Penentuan simpul terminal. Suatu simpul dijadikan simpul terminal jika max $\phi(s,t) \leq 0{,}006\ JKS(t_i)$ (Breiman, 1994).

   b. Penghentian pembentukan pohon regresi berdasarkan banyaknya amatan pada simpul akhir. Dalam penelitian ini banyak simpul akhir adalah 7

dengan banyak amatan pada simpul akhir pertama sampai simpul akhir 7 secara berurutan adalah 7. 5, 381, 56, 66, 5 dan 9.

c. Pemangkasan pohon regresi dengan menggunakan ukuran *cost complexity*. Ukuran *cost complexity* untuk setiap pemangkasan dapat dilihat pada lampiran 2.

d. Pemilihan pohon regresi terbaik dengan menggunakan sampel uji atau validasi silang.

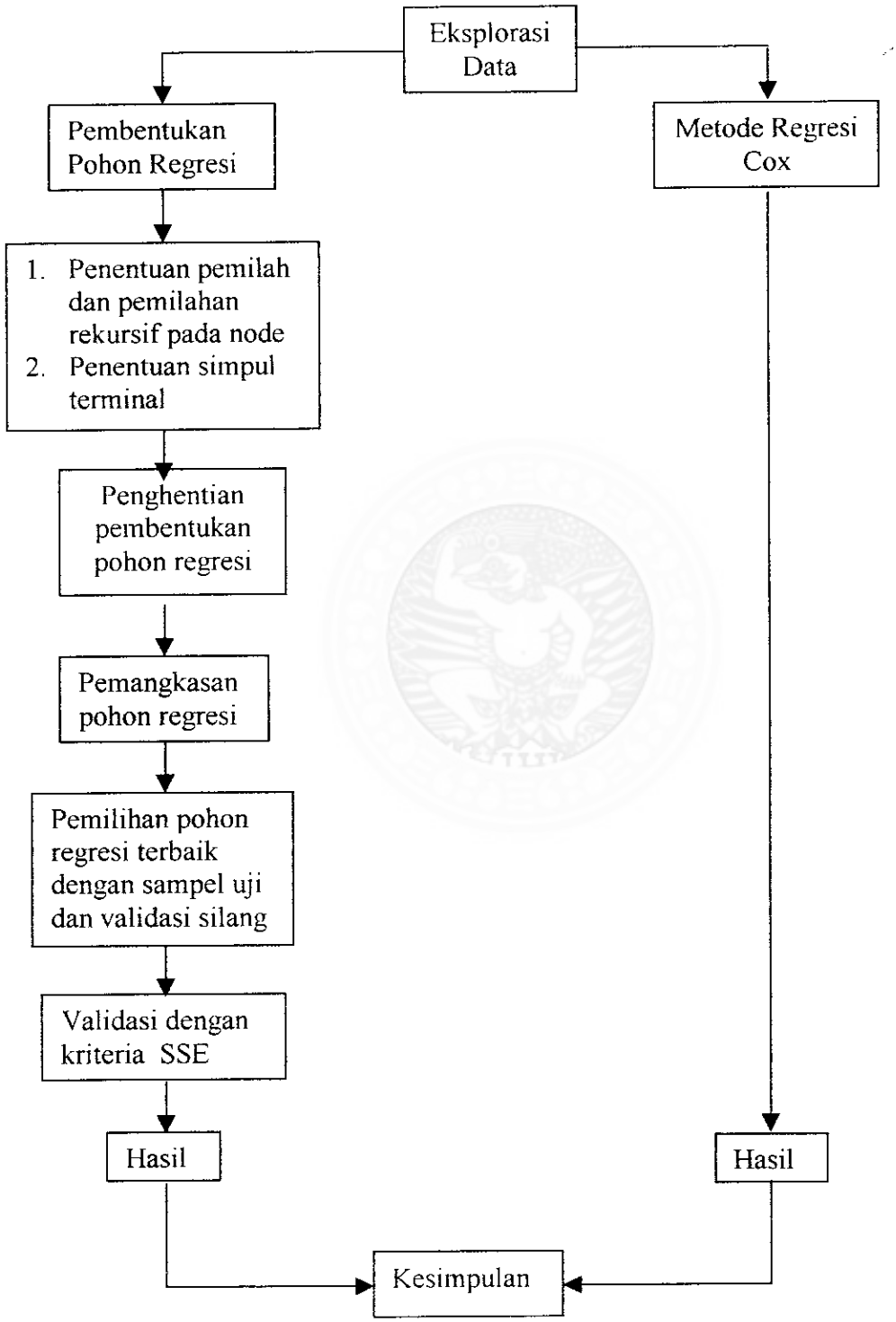
e. Menganalisis hasil.

3. Melakukan uji regresi Cox dengan menggunakan metode *forward stepwise (Likelihood Ratio)*.

4. Membandingkan hasil analisis faktor yang mempengaruhi lama perawatan penyakit diare pada di rumah sakit dengan menggunakan metode pohon regresi dan metode regresi Cox.

Proses analisis data tersebut dapat dilihat pada diagram di bawah ini.

Diagram Proses Analisis Data



BAB 4

HASIL PENELITIAN

BAB 4
HASIL PENELITIAN

4.1 Data Penelitian

Data tentang lama perawatan penyakit diare pada balita di rumah sakit Dr. Soetomo selama penelitian berjumlah 586 balita yang terdiri dari 57 (9,7%) balita dengan amatan lama perawatan yang tersensor dan 529 (90,3%) balita dengan amatan lama perawatan yang tidak tersensor. Balita tersebut terdiri dari 336 (57,3%) laki-laki dan 250 (42,7%) perempuan. Frekuensi tingkat dehidrasi yang dialami balita dapat dilihat pada tabel tabel 4.1.

Tabel 4.1. Frekuensi Tingkat Dehidrasi Balita

Tingkat Dehidrasi	Frekuensi (Orang)	Persentase (%)
Tidak ada	35	6,0
Ringan	166	28,3
Sedang	333	56,8
Berat	52	8,9
Jumlah	586	100,0

Pada tabel 4.1 di atas terlihat bahwa balita penderita diare yang dirawat di Rumah Sakit Dr. Soetomo, paling banyak mengalami dehidrasi sedang (56,8%) dan balita yang tidak mengalami dehidrasi sebesar 6,0%. Adapun frekuensi penyakit penyerta yang diderita balita penderita diare yang dirawat di rumah sakit dapat dilihat pada tabel 4.2 di bawah ini.

Tabel 4.2. Frekuensi Penyakit Penyerta yang Diderita Balita

Penyakit Penyerta	Frekuensi (orang)	Persentase (%)
Bronkopnemoni	28	4,8
Meningitis	2	0,3
Ensefalitis	6	1,0
Meningoensefalitis	1	0,2
Ensefalopati	22	3,8
Malnutrisi	71	12,1
Anemia	1	0,2
Marasmus	4	0,7
Kwarshiorkor	3	0,5
Marasmikwarshiorkor	2	0,3
Kejang Demam	25	4,3
Rhinopharyngitis	2	0,3
Lain-lain (bedah)	37	6,3
Tidak ada	382	65,2
Jumlah	586	100,0

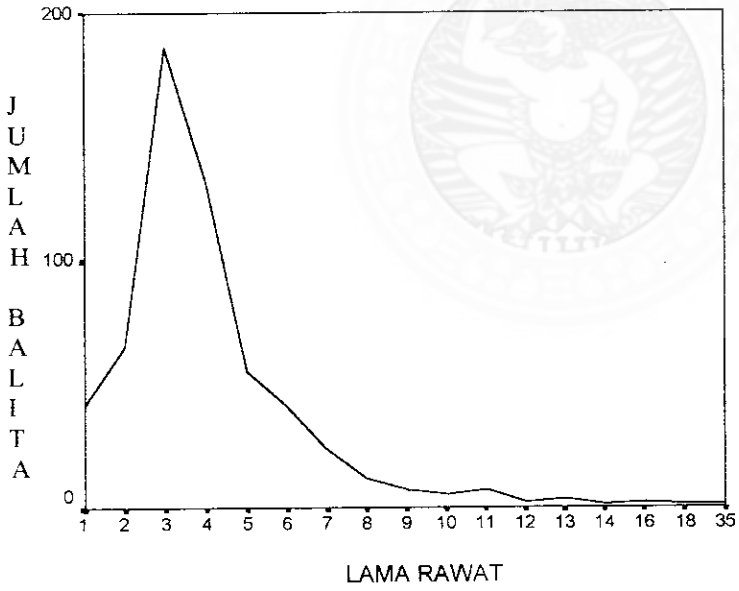
Berdasarkan tabel 4.2 terlihat bahwa 382 orang (65,2%) balita tidak mempunyai penyakit penyerta, sedangkan penyakit penyerta Meningoensefalitis dan Anemia masing-masing diderita oleh satu balita (0,2%). Gambaran umum data lama rawat, umur, lama diare dan status gizi balita dapat dilihat pada tabel 4.3.

Tabel 4.3. Gambaran Umum Data Lama Rawat, Umur, Lama Diare dan Status Gizi Balita.

Variabel	Min	Max	Range (Jangkauan)	Rata-rata	Simpangan baku
Lama Rawat (hari)	1,00	35,00	34,00	4,09	2,63
Umur (bulan)	0,06	60,00	59,94	11,55	10,13
Lama Diare (hari)	0,00	38,00	38,00	4,36	3,76
Status Gizi (NCHS-WHO)	8,18	181,31	173,13	82,13	15,99

Usia minimal balita penderita diare yang dirawat tersebut 0,06 bulan (1,8 hari) dan usia maksimal balita adalah 60 bulan dengan rata-rata umur balita sebesar 11,55 bulan. Balita yang menderita diare dibawa ke rumah sakit dengan rata-rata lama diare 4,36 hari, sedangkan status gizi balita penderita diare yang dirawat di rumah sakit berkisar dari 8,18 sampai dengan 181,3.

Variabel lama perawatan mempunyai data dengan modus sama dengan 3, median sama dengan 3,50 dan mean sama dengan 4,09. Data lama perawatan penyakit diare pada balita di rumah sakit mempunyai kurva model positif (sebaran menceng ke kanan) seperti terlihat pada gambar 4.1 di bawah ini.



Gambar 4.1. Diagram Garis Data Lama Perawatan Penyakit Diare pada Balita di Rumah Sakit Dr. Soetomo.

Tabel tabulasi silang antara penyakit penyerta, dehidrasi dan status gizi data balita penderita diare dapat dilihat pada tabel 4.4, sedangkan pada tabel 4.5 dapat dilihat tabel tabulasi silang antara penyakit penyerta, dehidrasi dan status gizi.

Tabel 4.4. Tabulasi Silang Penyakit Penyerta, Dehidrasi dan Status Gizi

		Stat Gizi ≤ 58,450		Total	**Stat Gizi > 58,450**		Total
		PP 2,6,9,10, 11,12,14	PP 1,3,4,5,7, 8,13		PP 2,6,9,10, 11,12,14	PP 1,3,4,5,7, 8,13	
Dehi-drasi 1	Jumlah	7	1	8	147	11	158
	Persentase	87,5	12,5	100	93,0	7,0	100
Dehi-drasi 2	Jumlah	10	8	18	323	79	402
	Persentase	55,6	44,4	100	80,3	19,7	100

Tabel 4.5. Tabulasi Silang Penyakit Penyerta, Dehidrasi dan Lama Diare

		Lama Diare ≤ 13,5		Total	**Lama Diare > 13,5**		Total
		PP 2,6,9,10, 11,12,14	PP 1,3,4,5,7, 8,13		PP 2,6,9,10, 11,12,14	PP 1,3,4,5,7, 8,13	
Dehi-drasi 1	Jumlah	151	12	163	3	0	3
	Persentase	92,6	7,4	100	100	0	100
Dehi-drasi 2	Jumlah	326	85	411	2	9	11
	Persentase	79,3	20,7	100	77,8	22,2	100

Keterangan :

Stat Gizi = Status Gizi

PP = Penyakit Penyerta, 1=Bronkopnemoni, 2=Meningitis, 3=Ensefalitis, 4=Meningoensefalitis, 5=Ensefalopati, 6=Malnutrisi, 7=Anemia 8=Marasmus, 9=Kwarshiorkor, 10=Marasmikwarshiorkor, 11=Kejang Demam, 12=Rhinopharyngitis, 13=Lain-lain (Bedah), 14=Tidak ada

Dehidrasi 1 = Dehidrasi ringan

Dehidrasi 2 = Dehidrasi sedang, berat dan tidak ada dehidrasi

Berdasarkan tabel 4.4 terlihat sebesar 87,5% balita penderita diare mengalami dehidrasi ringan dengan status gizi ≤ 58,450 dan mempunyai penyakit penyerta Meningitis, Malnutrisi, Kwarshiorkor, Marasmikwarshiorkor, Kejang Demam, Rhinopharyngitis dan tidak ada penyakit penyerta. Demikian juga pada tabel 4.5 terlihat sebesar 92,6% balita penderita diare mengalami dehidrasi ringan dengan lama diare kurang atau sama dengan 13,5 dan mempunyai penyakit penyerta Meningitis, Malnutrisi, Kwarshiorkor, Marasmikwarshiorkor, Kejang Demam, Rhinopharyngitis dan tidak ada penyakit penyerta.

Adapun persentase penyakit penyerta pada balita penderita diare yang mengalami dehidrasi ringan dapat dilihat pada tabel 4.6.

Tabel 4.6. Persentase Penyakit Penyerta pada Balita Penderita Diare dengan Dehidrasi Ringan

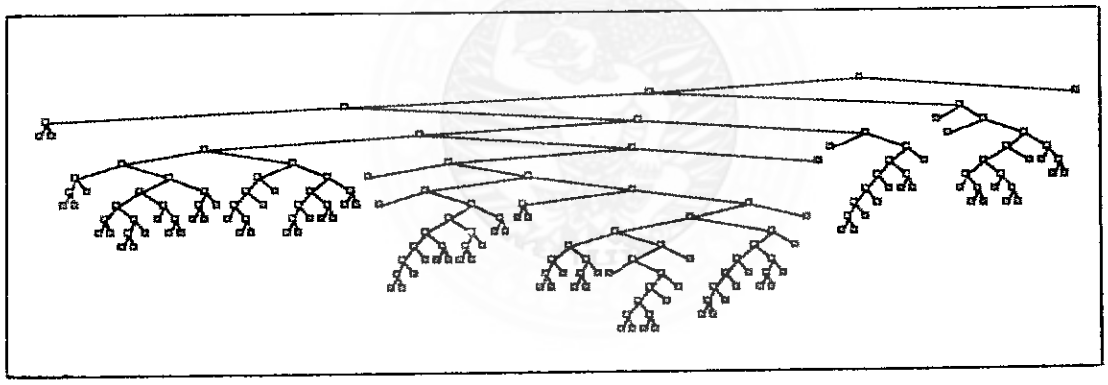
Penyakit Penyerta	**Persentase (%)**
Meningitis	0
Malnutrisi	26,8
Kwarshiorkor	0
Marasmikwarshiorkor	100
Kejang Demam	32,0
Rhinopharyngitis	100
Tidak ada	28,3

Dari tabel di atas terlihat balita penderita diare yang mengalami dehidrasi ringan paling banyak menderita penyakit penyerta Marasmikwarshiorkor dan Rhinopharyngitis.

4.2 Analisis Data

4.2.1 Analisis Metode Pohon Regresi

Proses pembentukan model pohon regresi pada data survival dilakukan berdasarkan tahap-tahap pembentukan yang telah dijelaskan pada bab terdahulu yaitu pertama-tama dilakukan pembentukan pohon regresi awal kemudian dari pohon regresi tersebut dilakukan pemilahan, penentuan ukuran pohon yang layak dan penentuan pohon regresi terbaik atau pohon yang optimal. Berdasarkan output CART 4.0 diperoleh pohon regresi awal pada gambar 4.2 dan pohon regresi optimal pada gambar 4.4.

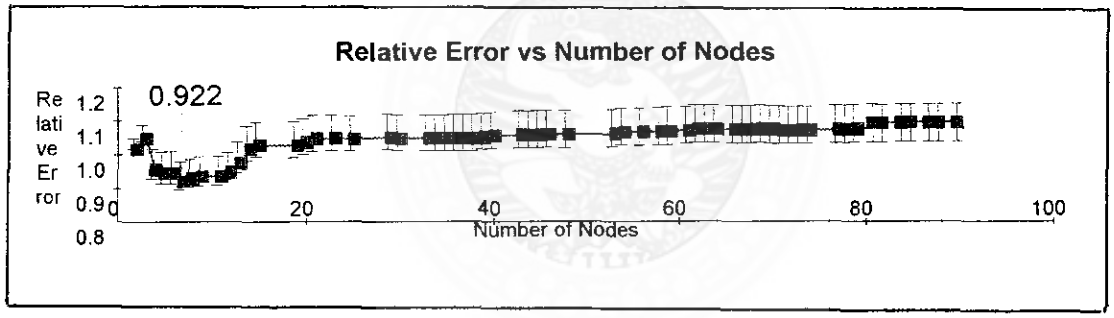


Gambar 4.2. Pohon Regresi Awal

Pada pembentukan pohon regresi awal terlihat pohon regresi tersebut mempunyai 90 simpul terminal (*terminal node*) di mana simpul pertama adalah variabel lama diare dengan nilai pemilah sebesar 13,5 hari. Hal ini menunjukkan bahwa lama diare mempunyai sumbangan relatif besar terhadap proses pembentukan pohon regresi untuk lama perawatan penyakit diare pada balita di Rumah Sakit Dr. Soetomo. Sesuai dengan posisi variabel lama diare yang

terletak pada simpul utama mempunyai nilai *Cross-Validate Relative Error* (R^{CV}) sebesar 1,080 ± 0,056, nilai *Resubstitution Relative Error* sebesar 0,531 dan *cost complexity* sebesar –1,000 (lihat lampiran 2).

Setelah dilakukan pemangkasan, diperoleh pohon regresi optimum. Untuk pohon regresi optimum diperoleh nilai R^{CV} sebesar 0,922 ± 0,041, nilai *Resubstitution Relative Error* sebesar 0,696 dan nilai *complexity* sebesar 45,497 (lampiran 2). Plot kesalahan relatif (*Relative Error*) dan jumlah simpul (*number of nodes*) untuk lama perawatan penyakit diare pada balita di rumah sakit dapat dilihat pada gambar 4.3.



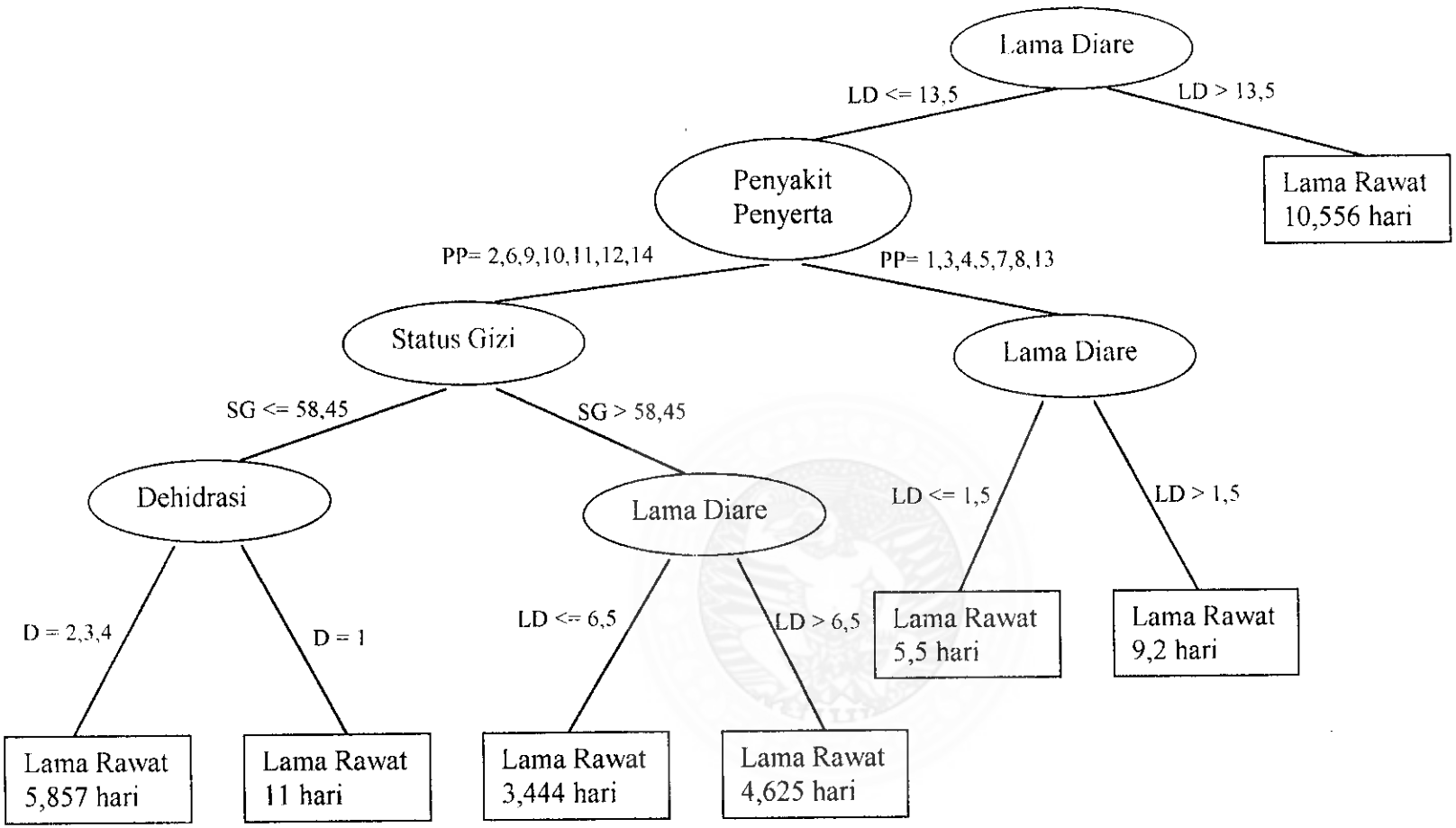
Gambar 4.3. Plot Kesalahan Relatif dan Jumlah Simpul Lama Perawatan Penyakit Diare pada Balita di Rumah Sakit Dr. Soetomo.

Model pohon regresi optimum untuk lama perawatan penyakit diare pada balita di Rumah Sakit Dr. Soetomo dapat dilihat pada gambar 4.4 (berdasarkan lampiran3). Pohon regresi optimum tersebut mempunyai mempunyai 1 simpul utama (*root node*), 5 simpul dalam (*internal* node) dan 7 simpul terminal (*terminal node*). Variabel independen yang masuk ke dalam simpul utama adalah variabel lama diare, sedangkan variabel yang masuk ke dalam simpul dalam adalah variabel penyakit penyerta, status gizi, dehidrasi dan variabel lama diare.

Hal ini berarti lama perawatan diare balita secara berurutan dipengaruhi oleh lama diare, penyakit penyerta, status gizi dan dehidrasi.

Pada gambar 4.4 terlihat juga bahwa pemilah pada simpul pertama adalah variabel lama diare dengan nilai sebesar 13,5. Variabel lama diare tersebut merupakan variabel independen yang dominan di antara variabel-variabel independen yang lain sehingga pohon regresi untuk lama perawatan balita yang menderita diare di Rumah Sakit Dr. Soetomo Surabaya banyak dijelaskan oleh variabel lama diare.

Simpul utama (simpul 1) dengan amatan sebanyak 529, rata-rata dan standar deviasi lama diare sebesar 4,104 hari dan 2,585. Simpul utama ini dipilah berdasarkan lama diare menjadi dua simpul yaitu simpul 2 dan simpul terminal 7. Pemilahan dilakukan terhadap amatan dengan lama diare balita kurang atau sama dengan 13,5 hari masuk ke simpul 2 dengan amatan sebanyak 520, sedangkan lama diare lebih dari 13,5 hari maka masuk ke dalam simpul terminal 7 dengan amatan sebanyak 9 yang mempunyai dugaan rata-rata lama perawatan diare balita tersebut adalah 10,556 hari.



Gambar 4.4. Pohon Regresi Optimal Lama Perawatan Penyakit Diare pada Balita di Rumah Sakit Dr. Soetomo

Keterangan :
LD = Lama Diare
PP = Penyakit Penyerta, 1 = Bronkopnemoni, 2 = Meningitis, 3 = Ensefalitis, 4 = Meningoensefalitis, 5 = Ensefalopati, 6 = Malnutrisi, 7 = Anemia, 8 = Marasmus, 9 = Kwarshiorkor, 10 = Marasmikwarshiorkor, 11 = Kejang Demam, 12 = Rhinopharyngitis, 13 = Lain-lain (Bedah), 14 = Tidak ada
SG = Status Gizi
D = Dehidrasi, 1 = Ringan, 2 = Sedang, 3 = Berat, 4 = Tidak ada

Variabel berikutnya yang berpengaruh pada simpul 2 adalah penyakit penyerta yang diderita balita. Simpul ini dipilah menjadi dua simpul. Jika balita menderita menderita penyakit Meningitis, Malnutrisi, Kwarshiorkor, Marasmikwarshiorkor, Kejang Demam, Rinopharyngitis dan tidak ada penyakit penyerta maka masuk ke simpul 3 dengan amatan sebanyak 449. Sebaliknya jika balita menderita penyakit Bronkopnemoni, Ensefalitis, Ensefalopati, Anemi, Marasmus dan penyakit lain-lain (bedah) masuk ke simpul 6 dengan amatan sebanyak 71. Pada simpul 6 terjadi pemilahan berdasarkan lama diare menjadi dua simpul. Jika lama diare kurang atau sama dengan 1,5 hari maka masuk ke simpul akhir 5 dengan amatan sebanyak 66, sedangkan jika lama diare lebih dari 1,5 hari maka masuk ke simpul akhir 6 dengan amatan sebanyak 5.

Variabel berikutnya yang berpengaruh pada simpul 3 adalah adalah status gizi balita. Simpul ini dipilah menjadi dua. Jika status gizi balita kurang atau sama dengan 58,45 maka masuk ke simpul 4 dengan amatan sebanyak 12 dan jika status gizi balita lebih dari 58,45 maka masuk ke simpul 5 dengan banyak amatan 437. Variabel berikutnya yang berpengaruh pada simpul 4 adalah dehidrasi. Simpul ini dipilah menjadi dua. Jika balita mengalami dehidrasi sedang, berat dan balita tidak mengalami dehidrasi maka masuk ke simpul akhir 1dengan amatan sebanyak 7, sedangkan jika balita mengalami dehidrasi ringan maka masuk ke simpul akhir 2 dengan amatan sebanyak 5. Pada simpul 5 terjadi pemilahan berdasarkan lama diare menjadi dua simpul. Jika lama diare kurang atau sama dengan 6,5 hari maka masuk ke simpul akhir 3 dengan amatan

sebanyak 381, sedangkan jika lama diare lebih dari 6,5 hari maka masuk ke simpul akhir 4 dengan amatan sebanyak 56.

4.2.2 Analisis Metode Regresi Cox

Untuk mengetahui lama perawatan penyakit diare pada balita di rumah sakit dengan menggunakan uji regresi Cox, ada tujuh variabel independen yang dianalisis yaitu umur, jenis kelamin, dehidrasi, penyakit penyerta, lama diare, muntah dan status gizi. Output uji regresi Cox dengan memasukkan semua variabel ke dalam model dapat dilihat pada lampiran 4.

Berdasarkan output uji regresi Cox tersebut, variabel independen yang masuk ke dalam model untuk $\alpha = 0,05$ adalah variabel penyakit penyerta, lama diare dan status gizi. Berhubung dari tujuh variabel yang dimasukkan ke dalam model, hanya tiga variabel yang signifikan maka dilakukan pengulangan uji regresi Cox dengan memasukkan ketiga variabel yang signifikan tadi ke dalam model. Output uji regresi Cox dapat dilihat pada lampiran 5.

Jika diperhatikan variabel yang masuk dalam model pada output uji regresi Cox (lampiran 4 dan lampiran 5), terlihat bahwa output regresi Cox dengan memasukkan semua variabel independen ke dalam model sama dengan output regresi Cox dengan memasukkan variabel yang signifikan saja ke dalam model. Dengan demikian dapat dikatakan bahwa untuk $\alpha = 0,05$ variabel yang berpengaruh di dalam model regresi Cox adalah variabel

penyakit penyerta 1 (Bronkopnemoni), penyakit penyerta 3 (Ensefalitis), penyakit penyerta 5 (Ensefalopati), penyakit penyerta 13 (Penyakit lain-lain/Bedah), lama diare dan status gizi. Taksiran nilai koefisien ($\hat{\beta}$) dan standar error (SE) dari variabel independen yang signifikan di dalam model regresi Cox dapat dilihat pada tabel 4.7.

Tabel 4.7. Taksiran Nilai Koefisien dan Standar *Error* dari Variabel Independen

Variabel	**Taksiran Nilai Koefisien ($\hat{\beta}$)**	**Standar Error (SE) $\hat{\beta}$**
Beronkopnemoni	- 0,726	0,222
Ensefalitis	-1,531	0,593
Enselopati	-1,064	0,309
Penyakit Lain-lain (Bedah)	-0,641	0,194
Lama Diare	-0,086	0,015
Status Gizi	0,008	0,003

Berdasarkan tabel 4.7 terlihat bahwa taksiran nilai koefisien ($\hat{\beta}$) untuk variabel penyakit penyerta Bronkopnemoni, Ensefalitis, Ensefalopati, Penyakit Lain-lain (Bedah) dan variabel lama diare bertanda negatif, sedangkan taksiran nilai koefisien ($\hat{\beta}$) variabel stutus gizi bertanda positif.

BAB 5

PEMBAHASAN

BAB 5

PEMBAHASAN

5.1 Pembahasan Metode Regresi Pohon

Berdasarkan hasil penelitian diperoleh karakteristik lama perawatan penyakit diare pada balita di Rumah Sakit Dr. Soetomo sebagai berikut :

1. Balita dengan lama diare sebelum dibawa ke rumah sakit lebih dari 13,5 hari maka rata-rata lama perawatannya di rumah sakit selama 10,556 hari dan simpangan baku sebesar 8,858.
2. Balita dengan lama diare sebelum dibawa ke rumah sakit kurang atau sama dengan 1,5 hari dan menderita penyakit Bronkopnemoni, Ensefalitis, Ensefalopati, Anemia, Marasmus dan penyakit lain-lain (bedah) maka rata-rata lama perawatannya 5,5 hari dan simpangan baku sebesar 2,835.
3. Balita yang menderita penyakit Bronkopnemoni, Ensefalitis, Ensefalopati, Anemia, Marasmus dan penyakit lain-lain (bedah) dengan 1,5 < lama diare ≤ 13,5 hari maka lama perawatan balita di rumah sakit adalah 9,2 hari simpangan baku 2,766.
4. Balita dengan lama diare sebelum dibawa ke rumah sakit kurang atau sama dengan 13,5 hari dan menderita penyakit Meningitis, Malnutrisi, Kwarshiorkor, Marasmikwarshiorkor, Kejang Demam, Rhinopharyngitis dan tidak mempunyai penyakit penyerta, status gizi kurang atau sama dengan 58,45 dengan dehidrasi sedang, berat dan tidak ada dehidrasi maka rata-rata

lama perawatan balita di rumah sakit selama 5,857 hari dan simpangan baku sebesar 2,900.

5. Lama diare balita sebelum dibawa ke rumah sakit kurang atau sama dengan 13,5 hari dan menderita penyakit Meningitis, Malnutrisi, Kwarshiorkor, Marasmikwarshiorkor, Kejang Demam, Rhinopharyngitis dan balita yang tidak mempunyai penyakit penyerta, gizi kurang atau sama dengan 58,45 dan mengalami dehidrasi ringan maka rata-rata lama perawatan balita di rumah sakit selama 11 hari dan simpangan baku sebesar 4,382.
6. Balita yang menderita penyakit Meningitis, Malnutrisi, Kwarshiorkor, Marasmikwarshiorkor, Kejang Demam, Rhinopharyngitis dan balita yang tidak mempunyai penyakit penyerta dengan status gizi lebih dari 58,45 dengan lama diare kurang atau sama dengan 6,5 hari maka rata-rata lama perawatan balita di rumah sakit adalah 3,444 hari dan simpangan baku 1,464.
7. Balita yang menderita penyakit Meningitis, Malnutrisi, Kwarshiorkor, Marasmikwarshiorkor, Kejang Demam, Rhinopharyngitis dan balita tersebut tidak mempunyai penyakit penyerta dengan status gizi lebih dari 58,45 dengan $6,5 <$ lama diare $\leq 13,5$ hari maka rata-rata lama perawatan balita di rumah sakit adalah 4,625 hari dan simpangan baku sebesar 1,625.

Berdasarkan karakteristik lama perawatan penyakit diare pada balita di rumah sakit terlihat bahwa lama perawatan balita yang menderita diare sebelum dibawa ke rumah sakit kurang atau sama dengan 13,5 hari dan menderita penyakit

Meningitis, Malnutrisi, Kwarshiorkor, Marasmikwarshiorkor, Kejang Demam, Rhinopharyngitis dan balita yang tidak mempunyai penyakit penyerta, gizi kurang atau sama dengan 58,45 dan mengalami dehidrasi ringan rata-rata lama perawatan balita tersebut di rumah sakit lebih lama dibandingkan dengan balita yang mengalami dehidrasi sedang, berat atau tidak ada dehidrasi. Jika diperhatikan tabel tabulasi silang pada lampiran 1 maka dapat dikatakan :

1. Berdasarkan tabel tabulasi silang penyakit penyerta, dehidrasi dan status gizi terlihat balita penderita diare yang menderita penyakit penyerta Meningitis, Malnutrisi, Kwarshiorkor, Maramikwarshiorkor, Kejang Demam, Rhinopharyngitis dan tidak ada penyakit penyerta, sebanyak 87,5% balita mengalami dehidrasi ringan dan sebanyak 55,6% balita mengalami dehidrasi sedang, berat dan tidak ada dehidrasi
2. Berdasarkan tabel tabulasi silang penyakit penyerta, dehidrasi dan lama diare terlihat balita penderita diare dengan penyakit penyerta Meningitis, Malnutrisi, Kwarshiorkor, Maramikwarshiorkor, Kejang Demam, Rhinopharyngitis dan tidak ada penyakit penyerta, terlihat sebanyak 92,6% balita mengalami dehidrasi ringan dan 79,3% balita menderita dehidrasi sedang, berat dan tidak ada dehidrasi.
3. Berdasarkan tabel tabulasi silang antara penyakit penyerta dan dehidrasi terlihat sebesar 100% balita menderita penderita diare juga menderita penyakit penyerta Marasmikwarshiorkor dan Rhinopharyngitis.

Balita penderita diare mengalami dehidrasi ringan lama perawatannya di rumah sakit lebih lama dibandingkan dengan balita penderita diare dengan dehidrasi sedang, berat atau tidak ada dehidrasi disebabkan karena penyakit penyerta Marasmikwarshiorkor dan Rhinopharyngitis yang diderita oleh balita tersebut.

5.2 Pembahasan Metode Regresi Cox

Berdasarkan output regresi Cox di atas, model regresi Cox untuk menduga lama perawatan diare balita ke-i di Rumah Sakit Dr. Soetomo adalah :

$$\widehat{h_i}(t) = \exp(-0{,}726\ \text{Bronkopnemoni}_{(i)} - 1{,}531\ \text{Ensefalitis}_{(i)} - 1{,}064\ \text{Ensefalopati}_{(i)} - 0{,}641\ \text{Penyakit Lain-lain (bedah)}_{(i)} - 0{,}086\ \text{Lama diare}_{(i)} + 0{,}008\ \text{Status gizi}_{(i)})\ \widehat{h_0}(t)$$

di mana :

Bronkopnemoni bernilai 1 jika balita penderita diare mempunyai penyakit penyerta Bronkopnemoni dan bernilai 0 jika balita penderita diare tidak menderita penyakit penyerta Bronkopnemoni. Ensefalitis bernilai 1 jika balita penderita diare menderita penyakit penyerta Ensefalitis dan bernilai 0 jika balita tidak menderita penyakit Ensefalitis. Ensefalopati bernilai 1 jika balita penderita diare menderita penyakit Ensefalopati dan bernilai 0 jika balita tidak menderita Ensefalopati. Penyakit lain-lain (bedah) bernilai 1 jika balita

penderita diare juga menderita penyakit lain-lain (bedah) dan bernilai 0 jika balita menderita diare tetapi tidak menderita penyakit lain-lain (bedah).

Berdasarkan persamaan di atas diperoleh taksiran *hazard rasio* untuk satu balita seperti terlihat pada tabel 5.1.

Tabel 5.1. Taksiran *Hazard Ratio* untuk Satu Balita.

Variabel Dalam Model	**Taksiran Koefisien**	**Taksiran *Hazard Ratio***
Bronkopnemoni	- 0,726	$e^{-0,726} = 0,484$
Ensefalitis	- 1,531	$e^{-1,531} = 0,216$
Ensefalopati	- 1,064	$e^{-1,064} = 0,345$
Penyakit Lain-lain (bedah)	- 0,641	$e^{-0,641} = 0,527$
Lama diare	- 0,086	$e^{-0,086} = 0,918$
Status gizi	0,008	$e^{0,008} = 1,008$

Dari tabel di atas terlihat bahwa Penyakit Bronkopnemoni, Ensefalitis, Ensefalopati, Penyakit lain-lain (bedah) dan lama diare mempunyai nilai taksiran *hazard ratio* lebih kecil dari 1. Hal ini menunjukkan bahwa balita penderita diare yang menderita penyakit Bronkopnemoni, Ensefalitis, Ensefalopati atau penyakit lain-lain (bedah) mempunyai lama perawatan yang lebih lama di rumah sakit. Begitu juga balita penderita diare dimana balita tersebut mempunyai waktu yang lama sebelum dibawa ke rumah sakit maka lama perawatan balita tersebut lebih lama dibandingkan dengan balita penderita diare yang lebih cepat dibawa ke rumah sakit. Selanjutnya nilai taksiran *hazard ratio* untuk status gizi balita lebih besar dari 1, berarti

semakin tinggi status gizi balita penderita diare semakin sedikit waktu yang diperlukan untuk perawatan balita tersebut di rumah sakit.

Adapun taksiran *baseline hazard function* $(\hat{h}_0(t))$, *survivor function* $(\hat{S}_0(t))$ dan *cumulative hazard function* $(\hat{H}_0(t))$ dapat dilihat pada tabel 5.2.

Tabel 5.2. Taksiran *Baseline Hazard Function, Survivor Function* dan *Cumulative Hazard Function* Lama Perawatan Balita Penderita Diare di Rumah Sakit.

Lama Perawatan	***Baseline Hazard Function***	***Survivor Function***	***Cumulative Hazard Function***
1	0,046	0,950	0,052
2	0,127	0,866	0,144
3	0,527	0,552	0,594
4	1,043	0,308	1,177
5	1,489	0,186	1,680
6	2,046	0,099	2,308
7	2,537	0,057	2,861
8	2,875	0,039	3,242
9	3,181	0,028	3,588
10	3,434	0,021	3,874
11	3,779	0,014	4,263
12	4,140	0,009	4,670
13	4,638	0,005	5,231
14	5,009	0,004	5,650
16	6,388	0,001	7,205
18	9,060	0,000	10,219
35	-	0,000	-

Berdasarkan tabel 5.2 di atas, terlihat *baseline hazard function* cenderung meningkat sesuai dengan peningkatan lama perawatan balita yang menderita diare di rumah sakit. Taksiran median waktu survival adalah waktu yang mempunyai taksiran fungsi survival kurang dari 0,5. Ini diperoleh dari balita penderita diare dengan lama perawatannya di rumah sakit selama 4 hari.

5.3 Komparasi Metode Pohon Regresi dengan Metode Regresi Cox

Berdasarkan uraian di atas, dengan menggunakan metode pohon regresi terdapat empat variabel independen yang berpengaruh terhadap lama perawatan penyakit diare pada balita di Rumah Sakit Dr. Soetomo. Variabel yang berpengaruh tersebut adalah lama diare sebelum balita dibawa ke rumah sakit, penyakit penyerta yang diderita oleh balita, status gizi balita dan tingkat dehidrasi yang dialami oleh balita tersebut. Dengan menggunakan metode regresi Cox terdapat tiga variabel independen yang berpengaruh terhadap lama perawatan penyakit diare pada balita di rumah sakit yaitu lama diare balita sebelum dibawa ke rumah sakit, penyakit penyerta yang diderita balita dan status gizi balita.

Selain dari itu, pada metode regresi pohon dapat diduga langsung berapa lama perawatan balita di rumah sakit, misalnya balita yang menderita penyakit Meningitis, Malnutrisi, Kwarshiorkor, Marasmikwarshiorkor, Kejang Demam, Rhinopharyngitis dan balita yang tidak mempunyai penyakit penyerta dengan status gizi lebih dari 58,45 dengan lama diare kurang atau sama dengan 6,5 hari maka rata-rata lama perawatan balita di rumah sakit adalah 3,444 hari sedangkan pada metode regresi Cox, lama perawatan ditaksir berdasarkan probabilitas yang dipengaruhi oleh *baseline hazard function.*

Adapun kelebihan metode regresi pohon adalah metode ini dapat menduga langsung berapa lama perawatan penyakit diare pada balita di rumah sakit seperti balita dengan lama diare sebelum dibawa ke rumah sakit lebih dari 13,5 hari maka lama perawatan balita di rumah sakit adalah selama 10,56 hari. Kelemahan

dari metode pohon regresi adalah metode ini tidak dapat menentukan kategori mana dari data yang sangat berpengaruh di dalam model jika datanya berskala nominal atau ordinal.

Kelebihan dari metode regresi Cox adalah dapat diketahui secara langsung tingkat signifikansi variabel yang berpengaruh di dalam model. Kelemahannya adalah tidak dapat secara langsung menentukan lama perawatan penyakit diare balita di rumah sakit.

Secara umum perbedaan hasil analisis metode pohon regresi dan metode regresi Cox dapat dilihat pada tabel 5.3.

Tabel 5.3. Perbedaan Hasil Analisis Metode Pohon Regresi dengan Metode Regresi Cox.

Metode Pohon Regresi	**Metode Regresi Cox**
1. Variabel yang masuk di dalam model adalah lama diare, penyakit penyerta, dehidrasi dan status gizi.	1. Variabel yang masuk di dalam model adalah penyakit penyerta, lama diare dan status gizi.
2. Dapat dengan langsung menentukan lama perawatan balita yang menderita diare di rumah sakit.	2. Tidak dapat dengan langsung menentukan lama perawatan balita yang menderita diare di rumah sakit.
3. Tidak dapat dilihat secara langsung tingkat signifikasi variabel yang masuk di dalam model.	3. Dapat dilihat secara langsung tingkat signifikansi variabel yang masuk di dalam model.
4. Interpretasi hasil lebih sederhana	4. Interpretasi hasil lebih rumit
5. Pengelompokan atau pemilahan terjadi secara otomatis.	5. Pengelompokan berdasarkan pengaruh tiap kategori.
6. Besar pengaruh tiap variabel independen tidak diketahui (p tidak ada)	6. Besar pengaruh tiap variabel independen ada (p ada)
7. Lebih tepat digunakan untuk data berskala nominal atau ordinal sebaiknya data mempunyai 2 (dua) kategori	

BAB 6

KESIMPULAN DAN SARAN

BAB 6

KESIMPULAN DAN SARAN

6.1 Kesimpulan

Penelitian ini dapat disimpulkan sebagai berikut :

1. Variabel independen yang berpengaruh terhadap lama perawatan penyakit diare pada balita di Rumah Sakit Dr. Soetomo dengan mengunakan metode pohon regresi adalah variabel lama diare, penyakit penyerta, status gizi dan dehidrasi.
2. Variabel independen yang berpengaruh terhadap lama perawatan penyakit diare pada balita di Rumah Sakit Dr. Soetomo dengan mengunakan metode regresi Cox adalah variabel penyakit penyerta, lama diare dan status gizi.
3. Terdapat perbedaan variabel yang berpengaruh terhadap lama perawatan penyakit diare balita di Rumah Sakit Dr. Soetomo dengan menggunakan metode pohon regresi dan metode regresi Cox.

6.2 Saran.

1. Jika ingin menduga lama perawatan penyakit diare pada balita di rumah sakit secara langsung dapat digunakan metode pohon regresi.
2. Jika ingin menduga lama perawatan penyakit balita di rumah sakit dengan menggunakan probabilitas dapat digunakan metode regresi Cox.
3. Jika ingin melihat pengaruh variabel independen terhadap variabel dependen di mana variabel independen mempunyai data berskala nominal atau ordinal

ADLN - Perpustakaan Universitas Airlangga

dengan kategori sama dengan 2 (dua) sebaiknya digunakan metode pohon regresi, tetapi jika ingin melihat pengaruh variabel independen terhadap variabel dependen di mana variabel independen mempunyai data berskala nominal atau ordinal dengan kategori lebih dari 2 (dua) sebaiknya digunakan metode regresi Cox.

DAFTAR PUSTAKA

ADLN - Perpustakaan Universitas Airlangga

MILIK
PERPUSTAKAAN
UNIVERSITAS AIRLANGGA
SURABAYA

DAFTAR PUSTAKA

Algifari., 2000. *Analisis Regresi. Teori, Kasus dan Solusi.* Edisi 2. BPFE. Yogyakarta.

Breiman, L, J. H. Friedman, R. A. Olshen, C. J. Stone., 1994. *Classification and Regression Trees.* Champman & Hall, London.

Collect, D., 1994. *Modeling Survival Data in Medicine Research.* Champman & Hall, London.

David, R. B & Anderson, J. K., 1989. Exponential Survival Trees. *Statist. Med.* 8 : 947-961.

Depkes RI, 2000. *Pedoman Pelaksanaan Program P2 Diare.* Ditjend P2M&PLP, Jakarta

Garpersz, V. 1995., *Teknik Analisis dalam Penelitian Percobaan.* Jilid 2. Tarsito, Bandung.

Hougaard P., 1986. Survival Models for Heterogenous Population Derived from Stable Distribution. *Biometrika* 73, No. 2 : 387 – 394.

Lee, E. T. 1980. *Statistical Methods for Survival Data Analysis.*Lifetime Learning Publications Belmont, California.

Krisna, I. K. B. & Supardi, S., 1996. Faktor-faktor resiko yang mempengaruhi Kejadian Diare Berdarah pada Balita di Kabupaten Klaten. *Berita Kedokteran. Masy.* XII (2) : 30 – 40.

ADLN - Perpustakaan Universitas Airlangga

Lawless, J. F., 1982. *Statistical Models and Methods for Lifetime Data.* John Willey & Sons Inc, New York.

Lubis, I. Z., 1992. Risk Faktors in Infatile Diarrhea (A Case Control Study). *Paediatrica Indosiana* 32 : 125 –134.

Miller, R. G. Jr., G. Gong, A. Munoz., 1981. *Survival Analysis.* John Wiley & Sons, New York.

Montgomery, D. C and E. A. Peek., 1992. *Introduction to Linear Regression Analysis. Second Edition.* John Willey & Sons Inc, New York.

Nelson W., 1982. *Applied Life Data Analysis.* John Willey & Sons Inc, New York.

Purwito, L., 1990. Hubungan antara Macam Air Menurut Sumbernya dengan Kejadian Diare di Kecamatan Sine Kabupaten Ngawi *(Tesis).* Universitas Airlangga, Surabaya.

Ranuh, N. G., 2001. *Kumpulan Kuliah Ilmu Kesehatan Anak.* Fakultas Kedoteran Universitas Hang Tuah Surabaya.

Segal, M. R., 1988. Regression Trees for Censored Data. *Biometrics* 44 : 35 – 47.

WHO., 1999. *Penatalaksanaan dan Pencegahan Diare.* Edisi III. WHO.

LAMPIRAN

Lampiran 1. Crosstabulasi Penyakit Penyerta, Dehidrasi dan Status Gizi data Balita Penderita Diare di Rumah Sakit

Case Processing Summary

	Cases					
	Valid		Missing		Total	
	N	Percent	N	Percent	N	Percent
PENYSERT * DEHIDRAS * STATGIZI	586	100,0%	0	,0%	586	100,0%
PENYSERT * DEHIDRAS * LADIARE	586	100,0%	0	,0%	586	100,0%

PENYSERT * DEHIDRAS * STATGIZI Crosstabulation

STATGIZI				DEHIDRAS ringan	DEHIDRAS sedang,berat, tidak ada	Total
< = 58.450	PENYSERT	Penyakit 2,6,9,10,11,12,14	Count	7	10	17
			% within PENYSERT	41,2%	58,8%	100,0%
			% within DEHIDRAS	87,5%	55,6%	65,4%
		Penyakit 1,3,4,5,7,8,13	Count	1	8	9
			% within PENYSERT	11,1%	88,9%	100,0%
			% within DEHIDRAS	12,5%	44,4%	34,6%
	Total		Count	8	18	26
			% within PENYSERT	30,8%	69,2%	100,0%
			% within DEHIDRAS	100,0%	100,0%	100,0%
> 58.450	PENYSERT	Penyakit 2,6,9,10,11,12,14	Count	147	323	470
			% within PENYSERT	31,3%	68,7%	100,0%
			% within DEHIDRAS	93,0%	80,3%	83,9%
		Penyakit 1,3,4,5,7,8,13	Count	11	79	90
			% within PENYSERT	12,2%	87,8%	100,0%
			% within DEHIDRAS	7,0%	19,7%	16,1%
	Total		Count	158	402	560
			% within PENYSERT	28,2%	71,8%	100,0%
			% within DEHIDRAS	100,0%	100,0%	100,0%

PENYSERT * DEHIDRAS * LADIARE Crosstabulation

LADIARE				DEHIDRAS ringan	DEHIDRAS sedang,berat, tidak ada	Total
<= 13.5	PENYSERT	Penyakit 2,6,9,10,11,12,14	Count	151	326	477
			% within PENYSERT	31,7%	68,3%	100.0%
			% within DEHIDRAS	92,6%	79,3%	83,1%
		Penyakit 1,3,4,5,7,8,13	Count	12	85	97
			% within PENYSERT	12,4%	87,6%	100,0%
			% within DEHIDRAS	7,4%	20,7%	16,9%
	Total		Count	163	411	574
			% within PENYSERT	28,4%	71,6%	100,0%
			% within DEHIDRAS	100,0%	100,0%	100,0%
> 13.5	PENYSERT	Penyakit 2,6,9,10,11,12,14	Count	3	7	10
			% within PENYSERT	30,0%	70,0%	100,0%
			% within DEHIDRAS	100,0%	77,8%	83,3%
		Penyakit 1,3,4,5,7,8,13	Count		2	2
			% within PENYSERT		100,0%	100,0%
			% within DEHIDRAS		22,2%	16.7%
	Total		Count	3	9	12
			% within PENYSERT	25,0%	75,0%	100,0%
			% within DEHIDRAS	100,0%	100,0%	100,0%

PENYSERT * DEHIDRAS Crosstabulation

			DEHIDRAS				Total
			ringan	sedang	berat	tidak	
PENYSERT	bronkopnemoni	Count	1	13	8	6	28
		% within PENYSERT	3,6%	46,4%	28,6%	21,4%	100,0%
		% within DEHIDRAS	,6%	3,9%	15,4%	17,1%	4,8%
	meningitis	Count		2			2
		% within PENYSERT		100,0%			100,0%
		% within DEHIDRAS		,6%			,3%
	ensefalitis	Count		1	4	1	6
		% within PENYSERT		16,7%	66,7%	16,7%	100,0%
		% within DEHIDRAS		,3%	7,7%	2,9%	1,0%
	meningoensefalitis	Count				1	1
		% within PENYSERT				100,0%	100,0%
		% within DEHIDRAS				2,9%	,2%
	ensefalopati	Count	3	6	13		22
		% within PENYSERT	13,6%	27,3%	59,1%		100,0%
		% within DEHIDRAS	1,8%	1,8%	25,0%		3,8%
	malnutrisi	Count	19	48	1	3	71
		% within PENYSERT	26,8%	67,6%	1,4%	4,2%	100,0%
		% within DEHIDRAS	11,4%	14,4%	1,9%	8,6%	12,1%
	anemia	Count		1			1
		% within PENYSERT		100,0%			100,0%
		% within DEHIDRAS		,3%			,2%
	marasmus	Count	1	2		1	4
		% within PENYSERT	25,0%	50,0%		25,0%	100,0%
		% within DEHIDRAS	,6%	,6%		2,9%	,7%
	kwarshiorkor	Count		2		1	3
		% within PENYSERT		66,7%		33,3%	100,0%
		% within DEHIDRAS		,6%		2,9%	,5%
	marasmikwarshiorkor	Count	2				2
		% within PENYSERT	100,0%				100,0%
		% within DEHIDRAS	1,2%				,3%
	kejang demam	Count	8	13		4	25
		% within PENYSERT	32,0%	52,0%		16,0%	100,0%
		% within DEHIDRAS	4,8%	3,9%		11,4%	4,3%
	rhinopharyngitis	Count	2				2
		% within PENYSERT	100,0%				100,0%
		% within DEHIDRAS	1,2%				,3%
	lain-lain (bedah)	Count	7	19	5	6	37
		% within PENYSERT	18,9%	51,4%	13,5%	16,2%	100,0%
		% within DEHIDRAS	4,2%	5,7%	9,6%	17,1%	6,3%
	tidak ada	Count	123	226	21	12	382
		% within PENYSERT	32,2%	59,2%	5,5%	3,1%	100,0%
		% within DEHIDRAS	74,1%	67,9%	40,4%	34,3%	65,2%
Total		Count	166	333	52	35	586
		% within PENYSERT	28,3%	56,8%	8,9%	6,0%	100,0%
		% within DEHIDRAS	100,0%	100,0%	100,0%	100,0%	100,0%

Lampiran 2. Jumlah pohon (*Tree Number*), Jumlah Simpul Akhir (*Terminal Node*), R^{CV}, Resubstitution Relative Error dan Nilai Kompleksitas (*Complexity*)

File: DATA PENELITIAN TESIS.SAV
Target Variable: LAMARWT
Predictor Variables: UMUR, LAMDIARE, STATGIZI, KELAMIN, DEHIDRAS, PENYSERT, MUNTAH

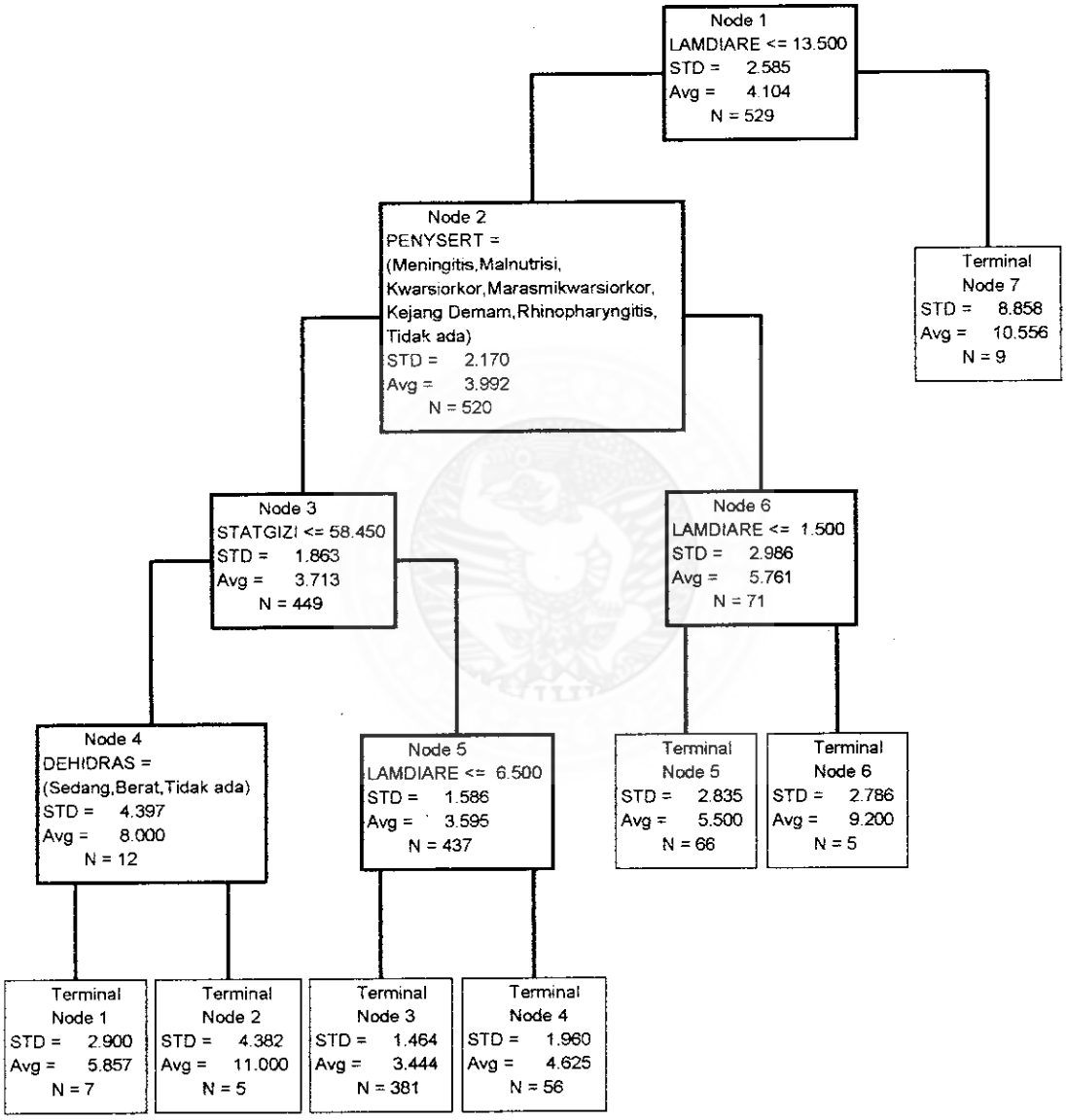
Tree Sequence

Tree Number	Terminal Nodes	Cross-Validated Relative Error	Resubstitution Relative Error	Complexity
1	90	1.080 ± 0.056	0.531	-1.000
2	88	1.081 ± 0.056	0.531	0.100
3	87	1.081 ± 0.056	0.531	0.148
4	85	1.081 ± 0.056	0.531	0.193
5	84	1.081 ± 0.056	0.531	0.325
6	82	1.079 ± 0.056	0.531	0.400
7	81	1.080 ± 0.056	0.532	0.529
8	79	1.078 ± 0.056	0.532	0.867
9	78	1.077 ± 0.056	0.532	0.900
10	77	1.076 ± 0.056	0.533	1.108
11	74	1.073 ± 0.056	0.534	1.158
12	73	1.074 ± 0.056	0.534	1.336
13	72	1.073 ± 0.056	0.534	1.350
14	71	1.074 ± 0.056	0.535	1.371
15	70	1.075 ± 0.056	0.535	1.556
16	69	1.078 ± 0.056	0.536	1.705
17	68	1.077 ± 0.056	0.536	1.894
18	67	1.075 ± 0.056	0.537	1.905
19	66	1.077 ± 0.056	0.537	1.985
20	64	1.077 ± 0.056	0.538	2.057
21	63	1.076 ± 0.056	0.539	2.083
22	62	1.077 ± 0.056	0.540	2.320
23	61	1.072 ± 0.055	0.540	2.795
24	59	1.071 ± 0.055	0.542	2.840
25	58	1.070 ± 0.055	0.543	2.844
26	56	1.069 ± 0.055	0.545	3.183
27	54	1.069 ± 0.055	0.547	3.467
28	53	1.064 ± 0.055	0.548	3.712
29	48	1.061 ± 0.055	0.553	4.046
30	46	1.061 ± 0.055	0.556	4.193
31	45	1.060 ± 0.055	0.557	4.821
32	44	1.059 ± 0.055	0.559	4.900
33	43	1.060 ± 0.055	0.560	5.429
34	40	1.057 ± 0.054	0.565	5.444
35	39	1.054 ± 0.054	0.567	6.349
36	38	1.049 ± 0.054	0.568	6.400
37	37	1.050 ± 0.054	0.570	6.577

Tree Number	Terminal Nodes	Cross-Validated Relative Error	Resubstitution Relative Error	Complexity
38	36	1.050 ± 0.054	0.572	7.421
39	35	1.050 ± 0.053	0.574	7.500
40	34	1.050 ± 0.053	0.577	7.915
41	33	1.050 ± 0.053	0.579	8.288
42	30	1.050 ± 0.053	0.586	8.302
43	29	1.054 ± 0.053	0.589	8.738
44	25	1.048 ± 0.053	0.599	9.510
45	23	1.049 ± 0.053	0.605	9.677
46	21	1.049 ± 0.053	0.611	10.765
47	20	1.039 ± 0.053	0.614	11.029
48	19	1.030 ± 0.053	0.618	13.279
49	15	1.031 ± 0.051	0.634	14.324
50	14	1.018 ± 0.051	0.638	15.248
51	13	0.976 ± 0.046	0.643	15.972
52	12	0.950 ± 0.041	0.649	22.500
53	11	0.937 ± 0.044	0.657	27.709
54	9	0.936 ± 0.043	0.674	30.726
55	8	0.931 ± 0.041	0.683	32.098
56**	7	0.922 ± 0.041	0.696	45.497
57	6	0.947 ± 0.047	0.714	63.630
58	5	0.947 ± 0.047	0.734	68.146
59	4	0.958 ± 0.045	0.755	77.143
60	3	1.050 ± 0.024	0.819	226.628
61	2	1.019 ± 0.016	0.892	257.094
62	1	1.001 ± 0.000802	1.000	381.101

* Minimum Cost
** Optimal

Lampiran 3. Pohon Optimal untuk Lama Perawatan Penyakit Diare pada Balita di Rumah Sakit Dr. Soetomo.



Lampiran 4. Output Regresi Cox dengan Memasukkan Variabel Umur, Jenis Kelamin, Dehidrasi, Penyakit Penyerta, Lama Diare, Muntah dan Status Gizi ke dalam Model

Cox Regression

Case Processing Summary

		N	Percent
Cases available in analysis	Event[a]	529	90,3%
	Censored	57	9,7%
	Total	586	100,0%
Cases dropped	Cases with missing values	0	,0%
	Cases with non-positive time	0	,0%
	Censored cases before the earliest event in a stratum	0	,0%
	Total	0	,0%
Total		586	100,0%

a. Dependent Variable: LAMARWT

Block 1: Method = Forward Stepwise (Likelihood Ratio)

Omnibus Tests of Model Coefficients[d,e]

Step	-2 Log Likelihood	Overall (score)			Change From Previous Step			Change From Previous Block		
		Chi-square	df	Sig.	Chi-square	df	Sig.	Chi-square	df	Sig.
1[a]	5861,037	27,301	1	,000	37,427	1	,000	37,427	1	,000
2[b]	5811,657	70,946	14	,000	49,380	13	,000	86,807	14	,000
3[c]	5804,060	80,839	15	,000	7,597	1	,006	94,404	15	,000

a. Variable(s) Entered at Step Number 1: LAMDIARE

b. Variable(s) Entered at Step Number 2: PENYSERT

c. Variable(s) Entered at Step Number 3: STATGIZI

d. Beginning Block Number 0, initial Log Likelihood function: -2 Log likelihood: 5898,464

e. Beginning Block Number 1, Method = Forward Stepwise (Likelihood Ratio)

Variables in the Equation

		B	SE	Wald	df	Sig.	Exp(B)
Step 1	LAMDIARE	-,078	,015	28,125	1	,000	,925
Step 2	PENYSERT			40,025	13	,000	
	PENYSERT(1)	-,709	,222	10,224	1	,001	,492
	PENYSERT(2)	-,524	,711	,544	1	,461	,592
	PENYSERT(3)	-1,357	,587	5,352	1	,021	,257
	PENYSERT(4)	-1,471	1,005	2,142	1	,143	,230
	PENYSERT(5)	-1,123	,308	13,296	1	,000	,325
	PENYSERT(6)	-,069	,136	,255	1	,613	,934
	PENYSERT(7)	-,800	1,003	,636	1	,425	,450
	PENYSERT(8)	-,457	,503	,826	1	,363	,633
	PENYSERT(9)	,022	,710	,001	1	,975	1,022
	PENYSERT(10)	,202	,710	,081	1	,776	1,224
	PENYSERT(11)	-,153	,226	,456	1	,499	,858
	PENYSERT(12)	,181	1,003	,032	1	,857	1,198
	PENYSERT(13)	-,705	,193	13,397	1	,000	,494
	LAMDIARE	-,085	,015	31,512	1	,000	,918
Step 3	PENYSERT			40,862	13	,000	
	PENYSERT(1)	-,726	,222	10,696	1	,001	,484
	PENYSERT(2)	-,652	,712	,839	1	,360	,521
	PENYSERT(3)	-1,531	,593	6,677	1	,010	,216
	PENYSERT(4)	-1,574	1,006	2,451	1	,117	,207
	PENYSERT(5)	-1,064	,309	11,877	1	,001	,345
	PENYSERT(6)	,074	,145	,261	1	,609	1,077
	PENYSERT(7)	-,710	1,004	,500	1	,479	,492
	PENYSERT(8)	-,309	,506	,373	1	,541	,734
	PENYSERT(9)	,078	,710	,012	1	,913	1,081
	PENYSERT(10)	,424	,714	,352	1	,553	1,527
	PENYSERT(11)	-,150	,226	,440	1	,507	,861
	PENYSERT(12)	,078	1,003	,006	1	,938	1,081
	PENYSERT(13)	-,641	,194	10,936	1	,001	,527
	LAMDIARE	-,086	,015	35,017	1	,000	,917
	STATGIZI	,008	,003	7,915	1	,005	1,008

Variables not in the Equation[a,b,c]

		Score	df	Sig.
Step 1	UMUR	,808	1	,369
	KELAMIN	,789	1	,374
	DEHIDRAS	8,256	3	,041
	DEHIDRAS(1)	1,121	1	,290
	DEHIDRAS(2)	1,148	1	,284
	DEHIDRAS(3)	6,321	1	,012
	PENYSERT	42,616	13	,000
	PENYSERT(1)	6,075	1	,014
	PENYSERT(2)	,224	1	,636
	PENYSERT(3)	4,020	1	,045
	PENYSERT(4)	1,627	1	,202
	PENYSERT(5)	10,440	1	,001
	PENYSERT(6)	,856	1	,355
	PENYSERT(7)	,364	1	,546
	PENYSERT(8)	,352	1	,553
	PENYSERT(9)	,066	1	,797
	PENYSERT(10)	,257	1	,612
	PENYSERT(11)	,018	1	,894
	PENYSERT(12)	,093	1	,760
	PENYSERT(13)	8,562	1	,003
	MUNTAH	3,314	1	,069
	STATGIZI	6,215	1	,013
Step 2	UMUR	,210	1	,647
	KELAMIN	,656	1	,418
	DEHIDRAS	,207	3	,976
	DEHIDRAS(1)	,124	1	,724
	DEHIDRAS(2)	,157	1	,692
	DEHIDRAS(3)	,056	1	,812
	MUNTAH	3,852	1	,050
	STATGIZI	7,848	1	,005
Step 3	UMUR	,825	1	,364
	KELAMIN	,454	1	,500
	DEHIDRAS	,115	3	,990
	DEHIDRAS(1)	,073	1	,787
	DEHIDRAS(2)	,115	1	,735
	DEHIDRAS(3)	,010	1	,919
	MUNTAH	3,650	1	,056

a. Residual Chi Square = 55,031 with 20 df Sig. = ,000

b. Residual Chi Square = 12,626 with 7 df Sig. = ,082

c. Residual Chi Square = 4,732 with 6 df Sig. = ,579

Model if Term Removed

Term Removed		Loss Chi-square	df	Sig.
Step 1	LAMDIARE	37,427	1	,000
Step 2	PENYSERT	49,380	13	,000
	LAMDIARE	41,821	1	,000
Step 3	PENYSERT	50,864	13	,000
	LAMDIARE	46,997	1	,000
	STATGIZI	7,597	1	,006

Covariate Means

	Mean
UMUR	11,551
KELAMIN	,573
DEHIDRAS(1)	,283
DEHIDRAS(2)	,568
DEHIDRAS(3)	,089
PENYSERT(1)	,048
PENYSERT(2)	,003
PENYSERT(3)	,010
PENYSERT(4)	,002
PENYSERT(5)	,038
PENYSERT(6)	,121
PENYSERT(7)	,002
PENYSERT(8)	,007
PENYSERT(9)	,005
PENYSERT(10)	,003
PENYSERT(11)	,043
PENYSERT(12)	,003
PENYSERT(13)	,063
LAMDIARE	4,358
MUNTAH	,703
STATGIZI	82,131

Lampiran 5. Output Regresi Cox dengan Memasukkan Variabel Penyakit Penyerta, Lama Diare dan Status Gizi ke dalam Model

Cox Regression

Case Processing Summary

		N	Percent
Cases available in analysis	Event[a]	529	90,3%
	Censored	57	9,7%
	Total	586	100,0%
Cases dropped	Cases with missing values	0	,0%
	Cases with non-positive time	0	,0%
	Censored cases before the earliest event in a stratum	0	,0%
	Total	0	,0%
Total		586	100,0%

a. Dependent Variable: LAMARWT

Block 1: Method = Forward Stepwise (Likelihood Ratio)

Omnibus Tests of Model Coefficients[d,e]

Step	-2 Log Likelihood	Overall (score)			Change From Previous Step			Change From Previous Block		
		Chi-square	df	Sig.	Chi-square	df	Sig.	Chi-square	df	Sig.
1[a]	5861,037	27,301	1	,000	37,427	1	,000	37,427	1	,000
2[b]	5811,657	70,946	14	,000	49,380	13	,000	86,807	14	,000
3[c]	5804,060	80,839	15	,000	7,597	1	,006	94,404	15	,000

a. Variable(s) Entered at Step Number 1: LAMDIARE

b. Variable(s) Entered at Step Number 2: PENYSERT

c. Variable(s) Entered at Step Number 3: STATGIZI

d. Beginning Block Number 0, initial Log Likelihood function: -2 Log likelihood: 5898,464

e. Beginning Block Number 1, Method = Forward Stepwise (Likelihood Ratio)

Variables in the Equation

		B	SE	Wald	df	Sig.	Exp(B)
Step 1	LAMDIARE	-,078	,015	28,125	1	,000	,925
Step 2	PENYSERT			40,025	13	,000	
	PENYSERT(1)	-,709	,222	10,224	1	,001	,492
	PENYSERT(2)	-,524	,711	,544	1	,461	,592
	PENYSERT(3)	-1,357	,587	5,352	1	,021	,257
	PENYSERT(4)	-1,471	1,005	2,142	1	,143	,230
	PENYSERT(5)	-1,123	,308	13,296	1	,000	,325
	PENYSERT(6)	-,069	,136	,255	1	,613	,934
	PENYSERT(7)	-,800	1,003	,636	1	,425	,450
	PENYSERT(8)	-,457	,503	,826	1	,363	,633
	PENYSERT(9)	,022	,710	,001	1	,975	1,022
	PENYSERT(10)	,202	,710	,081	1	,776	1,224
	PENYSERT(11)	-,153	,226	,456	1	,499	,858
	PENYSERT(12)	,181	1,003	,032	1	,857	1,198
	PENYSERT(13)	-,705	,193	13,397	1	,000	,494
	LAMDIARE	-,085	,015	31,512	1	,000	,918
Step 3	PENYSERT			40,862	13	,000	
	PENYSERT(1)	-,726	,222	10,696	1	,001	,484
	PENYSERT(2)	-,652	,712	,839	1	,360	,521
	PENYSERT(3)	-1,531	,593	6,677	1	,010	,216
	PENYSERT(4)	-1,574	1,006	2,451	1	,117	,207
	PENYSERT(5)	-1,064	,309	11,877	1	,001	,345
	PENYSERT(6)	,074	,145	,261	1	,609	1,077
	PENYSERT(7)	-,710	1,004	,500	1	,479	,492
	PENYSERT(8)	-,309	,506	,373	1	,541	,734
	PENYSERT(9)	,078	,710	,012	1	,913	1,081
	PENYSERT(10)	,424	,714	,352	1	,553	1,527
	PENYSERT(11)	-,150	,226	,440	1	,507	,861
	PENYSERT(12)	,078	1,003	,006	1	,938	1,081
	PENYSERT(13)	-,641	,194	10,936	1	,001	,527
	LAMDIARE	-,086	,015	35,017	1	,000	,917
	STATGIZI	,008	,003	7,915	1	,005	1,008

Variables not in the Equation[a,b]

		Score	df	Sig.
Step 1	PENYSERT	42,616	13	,000
	PENYSERT(1)	6,075	1	,014
	PENYSERT(2)	,224	1	,636
	PENYSERT(3)	4,020	1	,045
	PENYSERT(4)	1,627	1	,202
	PENYSERT(5)	10,440	1	,001
	PENYSERT(6)	,856	1	,355
	PENYSERT(7)	,364	1	,546
	PENYSERT(8)	,352	1	,553
	PENYSERT(9)	,066	1	,797
	PENYSERT(10)	,257	1	,612
	PENYSERT(11)	,018	1	,894
	PENYSERT(12)	,093	1	,760
	PENYSERT(13)	8,562	1	,003
	STATGIZI	6,215	1	,013
Step 2	STATGIZI	7,848	1	,005

a. Residual Chi Square = 50,409 with 14 df Sig. = ,000

b. Residual Chi Square = 7,848 with 1 df Sig. = ,005

Model if Term Removed

Term Removed		Loss Chi-square	df	Sig.
Step 1	LAMDIARE	37,427	1	,000
Step 2	PENYSERT	49,380	13	,000
	LAMDIARE	41,821	1	,000
Step 3	PENYSERT	50,864	13	,000
	LAMDIARE	46,997	1	,000
	STATGIZI	7,597	1	,006

Survival Table

Time	Baseline Cum Hazard	At mean of covariates		
		Survival	SE	Cum Hazard
1,000	,046	,950	,008	,052
2,000	,127	,866	,013	,144
3,000	,527	,552	,017	,594
4,000	1,043	,308	,015	1,177
5,000	1,489	,186	,014	1,680
6,000	2,046	,099	,010	2,308
7,000	2,537	,057	,008	2,861
8,000	2,875	,039	,007	3,242
9,000	3,181	,028	,006	3,588
10,000	3,434	,021	,005	3,874
11,000	3,779	,014	,004	4,263
12,000	4,140	,009	,004	4,670
13,000	4,638	,005	,003	5,231
14,000	5,009	,004	,002	5,650
16,000	6,388	,001	,001	7,205
18,000	9,060	,000	,000	10,219
35,000	.	,000	.	.

Covariate Means

	Mean
PENYSERT(1)	,048
PENYSERT(2)	,003
PENYSERT(3)	,010
PENYSERT(4)	,002
PENYSERT(5)	,038
PENYSERT(6)	,121
PENYSERT(7)	,002
PENYSERT(8)	,007
PENYSERT(9)	,005
PENYSERT(10)	,003
PENYSERT(11)	,043
PENYSERT(12)	,003
PENYSERT(13)	,063
LAMDIARE	4,358
STATGIZI	82,131